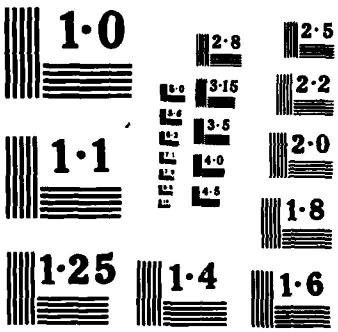
AD-A154 184   BAYESIAN FACTOR ANALYSIS(U) IOWA UNIV IOWA CITY   1/3
S I MAYEKAWA 23 MAR 85 TR-85-3-ONR N00014-83-C-0514

UNCLASSIFIED   F/G 12/1   NL


1·0
2·8
2·5
3·15
2·2
3·5
1·1
2·0
4·0
4·5
1·8
1·25
1·4
1·6

2

BAYESIAN FACTOR ANALYSIS

Shin-ichi Mayekawa

ONR Technical Report 85-3

CADA RESEARCH GROUP

THE UNIVERSITY OF IOWA

BAYESIAN FACTOR ANALYSIS

Shin-ichi Mayekawa

ONR Technical Report 85-3

March 23, 1985

This research was prepared under the Office of Naval Research Contract No. N00014-83-C-0514, Melvin R. Novick, Principal Investigator, The University of Iowa, Iowa City, Iowa.

unclassified March 23, 1985

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION unclassified	1b RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT approved for public release; distribution unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Melvin R. Novick 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION Office of Naval Research
6c ADDRESS (City, State, and ZIP Code) 356 Lindquist Center The University of Iowa Iowa City, IA 52242	7b ADDRESS (City, State, and ZIP Code) 536 South Clark Street Chicago, IL 60605
8a NAME OF FUNDING/SPONSORING ORGANIZATION Personnel & Training Res. Prog. 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-83-C-0514
8c. ADDRESS (City, State, and ZIP Code) Office of Naval Research (Code 458) Arlington, VA 22217	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO: PROJECT NO: NR150-521 TASK NO: WORK UNIT ACCESSION NO:

11 TITLE (Include Security Classification)

BAYESIAN FACTOR ANALYSIS Unclassified

12 PERSONAL AUTHOR(S)

Shin-ichi Mayekawa

13a TYPE OF REPORT	13b TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Technical Report 85-3	FROM 7/1/83 TO 3/23/85	1985 March 23	170

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB GROUP	Factor analysis, Bayesian estimation, EM algorithm.

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

A new Bayesian procedure for factor analysis is developed in which factor scores as well as factor loadings and error variances are treated as parameters of interest. The presentation is fully Bayesian in the sense that all the parameters have prior distributions and the posterior mode of a subset of the parameters is used as the point estimate.

The model is a standard one where the observations are expressed as the sum of the linear combination of factor scores, with factor loadings being the weights, and a normal error term. As the prior distribution the following exchangeable form is assumed:

A factor score vector for each observation has a common normal distribution.

A factor loading vector for each variable has a common normal distribution.

A error variance for each variable has a common inverted chi square distribution.

When the exchangeability of all the observations/variables is in question observations/variables may be divided into several subsets and the observations/variables within each subset may be treated as exchangeable. (over)

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Dr.Charles Davis	22b TELEPHONE (Include Area Code) (202) 696-4046 22c OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

unclassified

19. Abstract continued

Since the posterior marginal distribution of factor loadings and error variances can be expressed as the product of the covariance-based likelihood and the prior distributions of factor loadings and error variances the proposed method includes both the random and the fixed factor analysis models.

The mode of the hyperparameters is first derived from their posterior marginal distributions and conditional on those values the mode of error variance is derived from their posterior marginal distributions. Then, conditional of those estimates, the point estimate of factor scores and factor loadings are derived as the joint or the marginal mode of the posterior distribution of factor scores and factor loadings depending on the investigator's interest.

The marginalization is done via some variations of the EM algorithm and it is found that the different variations result in almost identical estimates. It is also found that the effect of the prior distribution of error variances is such that it reduces the number of local maxima. Finally, by specifying a priori zeros in the locational hyperparameters of factor loadings, a simple structure can be obtained without rotation.

Bayesian Factor Analysis*

Shin-ichi Mayekawa

The University of Iowa

Abstract

A new Bayesian procedure for factor analysis is developed in which factor scores as well as factor loadings and error variances are treated as parameters of interest. The presentation is fully Bayesian in the sense that all the parameters have prior distributions and the posterior mode of a subset of the parameters is used as the point estimate.

The model is a standard one where the observations are expresssed as the sum of the linear combination of factor scores, with factor loadings being the weights, and a normal error term. As the prior distribution the following exchangeable form is assumed:

- A factor score vector for each observation has a common normal distribution.
- A factor loading vector for each variable has a common normal distribution.
- A error variance for each variable has a common inverted chi square distribution.

When the exchangeability of all the observations/variables is in question observations/variables may be divided into several subsets and the observations/variables within each subset may be treated as exchangeable.

Since the posterior marginal distribution of factor loadings and error variances can be expressed as the product of the covariance-based likelikhood and the prior distributions of factor loadings and error variances the proposed method includes both the random and the fixed factor analysis models. —> cont keywords include ; 2.c 1473

The mode of the hyperparameters is first derived from their posterior marginal distributions and conditional on those values the mode of error variance is derived from their posterior marginal distributions. Then, conditional of those estimates, the point estimate of factor scores and factor loadings are derived as the joint or the marginal mode of the posterior distribution of factor scores and factor loadings depending on the investigator's interest.

The marginalization is done via some variations of the EM algorithm and it is found that the different variations result in almost identical estimates. It is also found that the effect of the prior distribution of error variances is such that it reduces the number of local maxima. Finally, by specifying a priori zeros in the locational hyperparameters of factor loadings, a simple structure can be obtained without rotation.

*Support for this research was provided under contract #N00014-83-C-0514 with the Personnel Training Branch of the Office of Naval Research, Melvin R. Novick, Principal Investigator.

Bayesian Factor Analysis

CHAPTER I

INTRODUCTION

Factor analysis is a multivariate statistical method used to explain the relationships among observed variables. Simply stated, standard factor analysis assumes that each observed variable is a weighted sum of two sets of random variables, namely, common factor scores and unique scores, all of which are unobservable. The purpose of the method is to estimate the weights, or factor loadings associated with each variable and to estimate the factor scores associated with each person. A typical application of the factor analysis method consists of the calculation of a correlation/dispersion matrix of the observed variables, which contains the sufficient statistics under the usual model, estimation of the weights, statistical testing of the model, and interpretation of the derived latent variables. Therefore, much of the literature of factor analysis is concerned with how to estimate factor loadings, how to test the model statistically, and how to find a meaningful interpretation of those latent variables.

Sometimes, however, it is desirable to go further and to estimate the values of those latent variables associated with each observation. For example, the vector unfolding model, which is often used to analyze the underlying structure of preference among a set of stimuli, has essentially the same model, Carroll(1972), or Bechtel(1976), and the scale values of each

stimulus must be known in practical applications. In applications of Spearman's theory of general intelligence it is the value of 'g' that is of interest in all applications. Also, the congeneric test model in classical test theory uses the same model as the factor analysis, Lord and Novick (1968), where the true score is represented by the factor scores. Therefore, if the individual true scores are needed they must be estimated. (See Chapter III for the detail.) However, as we shall see in Chapter II, these values cannot be determined uniquely under the standard factor analytic model because they are not treated as parameters of the model but remain as random variables even after the factor loadings are estimated. This general problem is known as the indeterminacy of factor scores, which, was probably first pointed out by Wilson (1928) and later elaborated by Guttman (1955). In this sense the standard factor analytic model was called the random factor analytic (RFA) model by McDonald (1979) and Anderson (1984).

The development of a model which enables us to estimate the values of those hypothetical concepts, namely, the factor scores, is not new. For example, Lawley (1942), Whittle(1952), Anderson and Rubin(1956), Joereskog(1963), McDonald and Burr(1967), and McDonald(1979b), have considered this problem, although usual textbooks do not discuss this model in detail,(see, for example, Harman's(1976, sec.2.3) treatment.) However, none of those methods were successful in the sense of providing unique estimates of factor scores. As we shall see later, this is due to the fact that the likelihood function of the fixed factor analytic (FFA) model, in which the factor scores are treated as parameters, is unbounded above, which implies nonexistence of maximum likelihood estimates.

The purpose of this thesis is to develop a method that enables the estimation of factor scores as parameters. Due to the nature of the problem the treatment is based on Bayesian fixed factor analysis. In Chapter II we first provide a brief review of the random factor analytic model and its classical and Bayesian estimation procedures. Then, in Chapter III, the classical fixed factor analytic model is introduced and its new Bayesian treatment is discussed in Chapter IV. It is shown that the method proposed is more general in the sense that both the RFA and the FFA models are included as special cases. In Chapter V, an evaluation of the new method is presented based on some real and artificial data sets.

CHAPTER II

REVIEW OF THE RANDOM FACTOR ANALYTIC (RFA) MODEL

Model

Factor analysis is one of several multivariate statistical methods studying the underlying relationships between observed variables. It assumes that each observed variable, y_j, $j=1,2,\ldots,p$, can be represented as a sum of three components:

$$(2.1.1) \quad y_j = m_j + \sum_{e=1}^{r} [\, f_e a_{je} \,] + u_j,$$

where the m_j is the overall mean of variable j, the f_e, $e=1,2,\ldots,r$, $r < p$, are latent (unobserved) variables called common factors, the a_{je}, $e=1,2,\ldots,r$, are the weights (called factor loadings) that link the e^{th} factor to variable j, and the u_j are other latent variables called unique factors for variable j. The number of latent common factors, r, is usually referred to the number of dimensions. Arranging p observed variables together, (2.1.1) can be written as

$$(2.1.2) \quad \underline{y} = \underline{m} + A\underline{f} + \underline{u},$$

where

$\underline{y} = [y_1, y_2, \ldots, y_p]'$, $p \times 1$,

$\underline{m} = [m_1, m_2, \ldots, m_p]'$, $p \times 1$,

$A = [\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_p]' = [\ a_{je}\ ]$, p x r,

$\underline{f} = [f_1, f_2, \ldots, f_r]'$, r x 1,

$\underline{u} = [u_1, u_2, \ldots, u_p]'$, p x 1.

Using $\underline{E}$, $\underline{D}$, and $\underline{C}$ to represent expectation, dispersion and covariance operators, respectively, the following specifications are typically made:

(2.1.3) $\underline{E}(\underline{f}) = \underline{0}$

(2.1.4) $\underline{D}(\underline{f}) = C_F$

where C_F is the r x r factor correlation matrix

with $\mathrm{diag}(C_F) = I_r$,

(2.1.5) $\underline{E}(\underline{u}) = \underline{0}$,

(2.1.6) $\underline{D}(\underline{u}) = D$,

where D is the p x p diagonal matrix consisting of d_j's,

(2.1.7) $\underline{C}(\underline{f}, \underline{u}) = 0$, (zero).

We call this model, following McDonald(1979b), the Random Factor Analysis (RFA) model, in the sense that $\underline{f}$ is treated as a random variable.

Under the RFA model we can deduce the following:

(2.1.8) $\underline{E}(\underline{y}) = \underline{m}$,

(2.1.9) $\underline{D}(\underline{y}) = \Omega = AC_FA' + D$,

$\underline{C}(\underline{y}, \underline{f}) = AC_F$,

(2.1.10) $\underline{E}(\underline{y}|\underline{f}) = \underline{m} + A\underline{f}$,

and

(2.1.11) $\underline{D}(\underline{y}|\underline{f}) = D.$

The last relation, (2.1.11), is sometimes called 'partial linear independence', Joereskog and Soerbom(1979, ch.1), or 'weak local independence', McDonald(1979a). The implication of (2.1.11) is that, given $\underline{f}$, the unique scores are not correlated with each other and that with respect to $\underline{f}$ the conditional variances of the unique scores are homoscedastic within each variable. It should be noted that the number of dimensions, r, is defined by (2.1.11). That is, r is the minimum number of common factors such that the conditional dispersion matrix of the observed variables given the common factors is diagonal. The only substantive assumptions of the RFA model are that $r < p$ and that the variables are conditionally independent and homoscedastic (2.1.11). Beyond that the model is simply a decomposition, cf. Lord and Novick, 1968, Ch. 24.

It can be shown that if the original variables are rescaled by p x p diagonal matrix V and a p x 1 vector $\underline{v}$ so that

(2.1.12) $\underline{y}^{*} = V(\underline{y} - \underline{v}),$

then the resulting variables have the mean and dispersion, respectively,

(2.1.13) $\underline{E}(\underline{y}^{*}) = V(\underline{m} - \underline{v}),$

$\underline{D}(\underline{y}^{*}) = V\Omega V.$

Therefore, the change of original scale results in the corresponding rescaling of the mean, factor loadings and unique variance, namely,

$\underline{m}^{*} = V(\underline{m}-\underline{v}),\quad A^{*} = VA,\quad \text{and}\quad D^{*} = VDV.$

When we estimate parameters from the observations, whether this property holds among the estimates depends on what method of estimation is used. It is known that maximum likelihood estimates and Bayesian mean, median, and modal estimates have this property. Some other estimation procedures do not have this property.

It is well known that the model (2.1.2) through (2.1.7) is not unique. Consider the transformation

(2.1.14) $\underline{f}^* = T'\underline{f}$,

where T is the r x r nonsingular matrix with $\mathrm{diag}(T'C_F T)=I_r$.

With this new latent variable we can rewrite the model as

(2.1.15) $\underline{y} = \underline{m} + B\underline{f}^* + \underline{u}$,

where $B = A\ (T')^{-1}$,

which has the same form as the original model.

With this new parametrization, we have

(2.1.16) $\underline{E}[\ \underline{f}^*\ ] = \underline{0}$,

(2.1.17) $\underline{D}[\ \underline{f}^*\ ] = T'C_F T$,

(2.1.18) $\underline{D}[\ \underline{y}\ ] = B\ T'C_F T\ B' + D$,

and

(2.1.19) $\underline{C}[\ \underline{y}, \underline{f}^*\ ] = B\ T'C_F T = AC_F\ T$.

This implies that, given one set of factors, we can always transform them into a desired form by appropriately choosing the transformation matrix T. This is known as rotational indeterminacy. For example, if the T matrix has the form

L4 Martin and McDonald's prior on D

Combining the prior distributions of A and C_F in L1 through L3 with the prior distribution of D of the form

the density of d_j is proportional to $\text{Exp}[(-1/2)v_j/d_j]$,

$j=1,2,\ldots,p$, independently,

where v_j's are prior constants,

we have L41, L42, and L43, say, respectively.

PR

For those elements of A of which we have strong prior information,

$\underline{a} \; : \; N(\, \underline{a}^* \, , \, G \,)$,

where $\underline{a}$ is the column roll-out of those elements of A.

For the rest of the elements of A, of which the information is vague, a locally uniform prior is assumed.

For each element of D, independently,

the density of d_j proportional to $1/d_j^h$,

is used.

$\underline{a}^*$, G, and h are the prior constants.

A and D are mutually independent.

A, C_F, and D are mutually independent.

L2 Non exchangeable factor loadings

For those elements of a_{je}'s not fixed, independently,

$a_{je} | s^2_{je}$: $N(a^*_{je} , s^2_{je})$,

$d_{je} w_{je} / s^2_{je}$: $\chi^2(d_{je})$.

For C_F and D the same priors as L1 is used.

a^*_{je}, d_{je}, w_{je}, for those elements of A not fixed,

are the prior constants.

A, C_F, and D are mutually independent.

L3 Noninformative prior

For those elements of A not fixed,

a_{je} : locally uniform.

For C_F, hierarchically,

$C_F^{-1} | R$: $W_r(R , g)$,

density of R is proportional to $|R|^{-(r+1)/2}$.

g is the prior constant.

2. Likelihood function

L1 through L4 and PR Sample dispersion matrix is assumed to have the Wishart density in (2.2.9). The likelihood L_1 in (2.2.10) is used.

KP Same as above with nonzero off diagonal element of D.

WO The data matrix Y is assumed to have the Normal density in (2.2.12). the likelihood L_2 will be used.

3. Prior distribution

L1 Exchangeable factor loadings

For those a_{je}'s which are not fixed, hierarchically,

$a_{je}|a^*,s^2 \ : \ N(a^*, s^2)$, i.i.d.,

$a^* \ : $ locally uniform,

$wd/s^2 \ : \ \chi^2(d)$.

For C_F,

$C_F^{-1} \ : \ W_r(R, g)$.

For each element of D, independently,

$h_j v_j / d_j \ : \ \chi^2(h_j)$.

d, w, R, g, and v_j's, and h_j's are prior constants.

5. Evaluate the posterior distribution in order to find some values of the parameters which best represent the posterior distribution, (location and dispersion.)

In the following section a step by step comparison of seven existing Bayesian methods will be presented. These methods are:

(L1) Lee's(1981) case 1.

(L2) Lee's(1981) case 2.

(L3) Lee's(1981) case 3.

(L4) Lee's(1981) case 4.

(PR) Press'(1982).

(KP) Kaufmann and Press'(1973).

(WO) Wong's(1980).

(Each method will be denoted by the abbreviation presented in the parenthesis.)

Because Martin and McDonald's(1975) method can be treated as a special case of Lee's method, it will not be considered explicitly.

Stepwise Comparison of Bayesian Methods

1. Parameters of interest

L1 through L4	A, C_F, and D with some of the elements of A fixed.
PR	A and D.
KP	A and D, with D not restricted to be a diagonal, r not restricted to be less than p.
WO	$\underline{m}$, A, and D.

approximation to the χ^2 distribution.

To test the hypothesis with some of the elements of A or C_F being fixed, say, to zero, a slightly different procedure is employed. This approach is called confirmatory factor analysis or the restricted model, [Joereskog (1969)]. Treating C_F as a parameter in the model, if the pattern of the values fixed is such that it enables the unique maximum of the likelihood, the minimization is performed without restrictions (2.2.15) and (2.2.16) with respect to D and those elements of A and C_F which are not fixed. If this is not the case the minimization will be performed subject to some additional condition which guarantees a unique solution. In either case, some adjustment of the degrees of freedom is necessary. Essentially, d.f. is equal to

(2.3.4) $\quad p(p+1)/2$ – number of free parameters to be estimated

+ number of independent restrictions on the parameters.

For further details, see Joereskog(1969).

The Review of Bayesian Estimation Methods in RFA Model

In this section a review of existing Bayesian estimation methods in the RFA model is presented. In general, Bayesian estimation proceeds as follows:

1. Identify the parameters of interest.
2. Specify the conditional distribution of data given parameters, namely, the likelihood function.
3. Specify the form of the prior distribution of the parameters.
4. Find the posterior distribution of the parameters given the data.

for the resulting normal equations. Since the grand mean $\underline{m}$ is treated as known in their method, the result is the same as the MLE based on the L_2 likelihood.

Goodness of Fit Test in RFA Model

With large N, it is known, Joereskog (1967), that minus twice the log likelihood ratio

$$(2.3.1) \quad LLR = -2(\ (\ L_1(A^+,D^+|S)\ )$$

$$- \ln(\ |S|^{-(N-1)/2}Exp[-(N-1)/2)tr(S^{-1}S)]\ )$$

$$= (N-1)(\ F_1(A^+,D^+) - \ln(|S|) - p\ )$$

is distributed as

$$LLR \ :\ X^2(\ (\ (p-r)^2 - (p+r)\ )/2\ ).$$

The inside of the second logarithm term is the value of L_1 evaluated at $\Omega = (N/(N-1))S$. Therefore, the hypothesis

$$(2.3.2) \quad H_r \ :\quad \Omega = AC_FA' + D \quad \text{with specified } r,$$

can be tested against the alternative hypothesis

$$(2.3.3) \quad H_0 \ :\quad \Omega \text{ is positive definite symmetric,}$$

with approximate significance level alpha, by comparing the value of LLR to the (1-alpha) percentile point of the X^2 distribution, i.e., if LLR exceeds the percentage point, H_r will be rejected. Bartlet(1951) has suggested that the use of N-1-(2p+5)/N-2r/3 in place of N-1 in (2.3.1) in order to obtain a better

matrix, thus, the Ω matrix, must be calculated by (2.2.23) each time D is updated since we are treating F_2 as the function of D only. The algorithm usually converges rapidly to a local minimum. It is known, however, that the minimum often lies on the region where some of the error variances are zero, i.e., a Heywood case.

Another approach to the ML estimation of the RFA model is found in Rubin and Thayer (1982) where the EM algorithm is used. (See Chapter IV for the explanation of the EM algorithm.) Their solution is the iteration of the following two steps:

The E-step:

(2.2.26) $W = (D + AC_FA')^{-1} AC_F$,

and

(2.2.27) $Q = (W' S W + V)^{-1}$,

where

(2.2.28) $V = C_F - C_FA(D+AC_FA')^{-1}AC_F$.

The M-step:

(2.2.28) $A = S W Q$,

and

(2.2.29) $D = \mathrm{diag}(S - SWQW'S)$.

The method is derived by writing the L_3 likelihood discussed in Chapter III in terms of the sufficient statistics, S, (1/N)Y'F, and (1/N)F'F, replacing the corresponding quantities by their conditional expectations given Y, namely, S, SW, and Q^{-1}, differentiating the result with respect to A and D, and solving

should be taken. Also, from (2.2.18) we have, given A,

(2.2.24) $D = \mathrm{diag}(\ S - AC_F A'\ )$.

Therefore, the iteration of (2.2.23) and (2.2.24) gives the desired minimum of F_2 when the process converges. However, this method given in Lawley and Maxwell (1963) is known to be very slow to converge. Therefore, the following method is usually used [Lawley and Maxwell (1971) or Joereskog (1967)]. Noticing the fact that the conditional minimum of F_2 with respect to A given D is obtained analytically by (2.2.23) we may consider F_2 as a function of D only. That is, the derivative of F_2 with respect to D can be evaluated as

(2.2.25) $d\, F_2/d\, d_j = \partial F_2/\partial d_j + \mathrm{tr}([\partial F_2/\partial A]'[\partial A/\partial d_j])$,

$$j = 1,2,\ldots,\ p.$$

However, the second term vanishes at the point where (2.2.23) is satisfied since $\partial F_2/\partial\, A$ is zero if the A matrix given in (2.2.23) is used to evaluate it. Therefore, the derivative of F_2 with respect to D, when it is regarded as a function of D only, is given by (2.2.18) with the A matrix, thus the Ω matrix, defined by (2.2.23). Given the derivatives, the minimization of F_2 can be performed via several existing numerical methods which do not require the information provided by the second derivative of F_2 with respect to D. For example, the Fletcher—Powell method, which is advocated by Joereskog (1967), determines the direction of search for the minimun by approximating the inverse of the second derivative matrix using the information provided by the first derivatives only. It sould be noted that when the evaluation of F_2 is necessary, say, in the cubic search along the direction determined, the A

solution based on F_2 is as follows. The partial derivatives of F_2 with respect to A and D are:

(2.2.17) $\partial F_2/\partial A = 2(\ \Omega^{-1}A-\Omega^{-1}S\Omega^{-1}A\ )$,

and

(2.2.18) $\partial F_2/\partial D = \mathrm{diag}(\ \Omega^{-1}-\Omega^{-1}S\Omega^{-1}\ )$.

(See the Appendix for matrix differentiation.)

By setting (2.2.17) equal to zero we have

(2.2.19) $S\Omega^{-1}A = A$.

But, by Lawley's trick in the Appendix, it can be reexpressed as

(2.2.20) $SD^{-1}A(I_r+A'D^{-1}A)^{-1} = A$,

thus,

(2.2.21) $SD^{-1}\ A = A\ (I_r+A'D^{-1}A)$,

or,

(2.2.22) $D^{-1/2}SD^{-1/2}\ D^{-1/2}A = D^{-1/2}A\ (I_r+A'D^{-1}A)$.

This form shows, with the restriction (2.2.15), that $(I_r+A'D^{-1}A)$ is the eigen values of $D^{-1/2}SD^{-1/2}$ and that $D^{-1/2}A$ is the associated eigen vectors. That is, given D, the A matrix which minimizes F_2 can be written as

(2.2.23) $A = D^{1/2}Q(L+I_r)^{1/2}$,

where L is the eigen value matrix of $D^{-1/2}SD^{-1/2}$ and Q is the associated orthonormal eigen vectors.

It can be shown that in order to minimize F_2 the r largest eigen values

$$\times \operatorname{Exp}[(-1/2)\operatorname{tr}((Y-\underline{1}_N\underline{m}')\,\Omega^{-1}(Y-\underline{1}_N\underline{m}')')]$$

$$= |\Omega|^{(-N/2)} \times \operatorname{Exp}[(-N/2)\operatorname{tr}(\Omega^{-1}S)]$$

$$\times \operatorname{Exp}[(-N/2)(\underline{y}_. - \underline{m})'\Omega^{-1}(\underline{y}_. - \underline{m})],$$

where $\underline{y}_. = (1/N)\sum_{i=1}^{N}[\ \underline{y}_i\ ]$.

Noticing that regardless of the value of Ω, L_2 is maximized at

(2.2.13) $\underline{m}^{+} = \underline{y}_.$,

the MLE of A and D can be found by minimizing the monotone decreasing function of L_2,

(2.2.14) $F_2(A,D) = \ln|\Omega| + \operatorname{tr}(\Omega^{-1}S)$.

Because F_2 is not a monotone function of F_1 the resulting solution, when N is not so large, is different.

In both methods, in order to remove the rotational indeterminacy the restrictions

(2.2.15) $A'D^{-1}A$ is diagonal,

and

(2.2.16) $C_F = I_r$,

are enforced. If the factors are assumed to be correlated, the correlated factor will be found by finding the transformation matrix P described in (2.1.14) through (2.1.20).

Several numerical methods are available, for example, see Lawley and Maxwell(1963,1971) or Joereskog(1967,1977). The outline of the

(2.2.6) $S = (1/N)Y'JY$,

where

(2.2.7) $Y = [\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_N]'$, $N \times p$,

and

(2.2.8) $J = I_N - (1/N)\underline{1}_N\underline{1}_N'$,

has the Wishart distribution, Press(1972),

(2.2.9) $N \times S \;:\; W_p(\, AC_FA' + D \,,\, N-1 \,)$,

where $W_p(\, C \,,\, df \,)$ denotes the Wishart distribution with the expectation $df \times C$ and the degrees of freedom, df.

The likelihood of A and D is proportional to

(2.2.10) $L_1(A,D|S) \;=\!=\; |\Omega|^{-(N-1)/2}\mathrm{Exp}[-(N/2)\mathrm{tr}(\Omega^{-1}S)]$,

where the symbol $=\!=$ denotes the proportionality.

Therefore, in this formulation, the statistical estimation of the original EFA parameters reduces to the estimation of the covariance structure shown in (2.1.9) under the Wishart probability model. (Usually, $\underline{m}$ is estimated by the sample overall mean.) The maximum likelihood estimates (MLE) of A and D are found by minimizing a monotone decreasing function of L_1

(2.2.11) $F_1(A,D) = \ln|\Omega| + (N/(N-1))\mathrm{tr}(\Omega^{-1}S)$.

Although this is the most commonly used ML solution there is another ML estimation procedure based on the original density of Y described in Anderson and Rubin(1956) and Mardia, et., al.(1979). The likelihood of $\underline{m}$, A, and D given Y can be written as

(2.2.12) $L_2(\underline{m},A,D|Y) \;=\!=\; |\Omega|^{(-N/2)}$

where Ω is defined in (2.1.9),

$N_p(\underline{m}, \Omega)$ denotes the p-variate normal distribution

with the mean vector $\underline{m}$ and the dispersion matrix Ω,

and the symbol : is used to denote the distributional law.

By applying Rao's (1973) formulae 8a.3(1) and 8a.2.9

it can be shown that if we replace (2.1.7) by its stronger form,

(2.2.2) $\underline{f}$ and $\underline{u}$ are statistically independent,

then (2.2.1), together with (2.1.2) through (2.1.6), imply,

(2.2.3) $\underline{f} \; : \; N_r(\underline{0}, C_F)$

and

(2.2.4) $\underline{u} \; : \; N_p(\underline{0}, D)$.

Therefore

(2.2.5) $\underline{y} \mid \underline{f} \; : \; N_p(\underline{m}+A\underline{f}, D)$.

Because of normality, (2.2.5) implies the local independence of $\underline{y}$ given $\underline{f}$ in the sense of Lord and Novick(1968,ch.24). However, it is not true that the normality of $\underline{y}$ and the weak local independence in (2.1.11) imply the strong local independence in (2.2.5) and the independence of $\underline{f}$ and $\underline{u}$ in (2.2.2). Conversely, in order to derive the normality of $\underline{y}$ in (2.2.1), we can start with (2.1.2), the independence of $\underline{f}$ and $\underline{u}$ in (2.2.2), and their normality in (2.2.3), and (2.2.4).

Having observed N independent observations, $\underline{y}_i$, i=1,2,...,N, it follows that the sample dispersion matrix

random, and, therefore, cannot be determined. Some interesting discussions of this point from a Bayesian point of view are found in Bartholomew (1981) and Mardia, Kent, and Bibby (1979).

Historically, the RFA model is stated in terms of the correlation matrix, that is, using $[\mathrm{diag}(\Omega)]^{-1/2}$ and $\underline{m}$ as the scaling constants V and $\underline{v}$, respectively, in (2.1.12), (2.1.9) can be written as

$$(2.1.22) \qquad \underline{D}(\underline{y}^*) = \mathrm{diag}(\Omega)^{-1/2}\, \Omega\, \mathrm{diag}(\Omega)^{-1/2} = R, \text{ say,}$$

where R is the population correlation matrix of the observed variables. However, instead of the population mean and standard deviation, which are not observable, if we use the sample analogues of those quantities as the scaling constants, (2.1.22) is no longer correct since $\underline{y}^*$ is not a linear transformation of $\underline{y}$. Similarly, if we treat those estimates as fixed scaling constants, (2.1.22) is not correct either since in this case R is not the population correlation matrix but simply a rescaled covariance matrix. Therefore throughout this paper we do not refer to the correlation matrix.

Maximum Likelihood Estimation in RFA

In this section we briefly review the Maximum Likelihood (ML) estimation procedure of the RFA parameters, namely $\underline{m}$, A, and D. In order to perform ML estimation it is necessary to introduce the following full distributional assumption:

$$(2.2.1) \qquad \underline{y} : N_p(\underline{m}, \Omega)$$

(2.1.20) $T = P\,M^{-1/2}\,Q\,M^{1/2}\,P'$,

where $C_F = P\,M\,P'$ is the normalized eigen decomposition of C_F,

and Q is any r x r orthonormal matrix,

the new variable $\underline{f}^*$ has the same correlation matrix C_F. If the T matrix has the form

(2.1.21) $T = P\,M^{-1/2}\,Q$,

the new variables become orthogonal, i.,e., $\underline{D}[\,\underline{f}^*\,] = I_r$.

It should be also noted that under the RFA model $\underline{f}$ is treated as a latent random variable and therefore cannot be estimated in the usual statistical sense, i.e., after $\underline{m}$, A, and D have been estimated $\underline{f}$ and $\underline{u}$ still remain as random *quantities. The so called 'problem of factor score* indeterminacy' stems partly from this fact. Usually, in order to 'estimate' the value of $\underline{f}$ associated with each object, i=1,2,...,N, some arbitrary least squares criterion is introduced and the value which minimizes the criterion, given the estimate of A and D, is sought. Without these additional criteria it is known, Guttman(1955), that, as a linear combination of $\underline{y}$, the factor scores cannot be uniquely determined even if we have estimated A and D uniquely. (See the Appendix for the derivation of the Guttman/Ketselman formula.) That is, given A and D, we can construct an infinite number of sets of $\underline{f}$ and $\underline{u}$ which satisfy the model (2.1.2) through (2.1.7). Shiba (1969) has identified at least sixteen methods with different criteria. Simply stated, given the particular A matrix and conditioning on the observed value of $\underline{y}$, $\underline{f}$ and $\underline{u}$ in (2.1.2) remain

KP

In this model, instead of using the (A,D) parametrization, the prior density of (A,Ω) is assumed to be proportional to

$$f(A,\Omega) \propto \mathrm{Exp}[(-1/2)\mathrm{tr}(A-A_2)V_2(A-A_2)'G)]$$

$$\times\ |\Omega|^{-(h+p+1)/2}\mathrm{Exp}[(-1/2)\mathrm{tr}(\Omega^{-1}Q(A))],$$

where $Q(A) = S_0 + (A-A_1)V_1(A-A_1)'$,

and h, V_1, V_2, A_1, A_2, and S_0 are the prior constants.

This form of density implies that given Ω, A is a truncated matrix normal distribution, and that given A, Ω is a truncated Wishart distribution. The truncations are done so that $D = \Omega - AA'$ is positive definite.

WO

For $\underline{m}$,

$$\underline{m}\ :\ N_p(\ \underline{0}\ ,\ \Omega_m\ ),\ \Omega_m^{-1} \rightarrow 0\ \text{(zero)}.$$

For each element of A, independently,

$$a_{je}\ :\ N(\ 0\ ,\ s^2_e\ ),\ e=1,2,\ldots,r.$$

Because the main purpose of Wong's study is the marginal maximum likelihood estimation of D, no prior distribution of D is used.

$\underline{m}$ and A are mutually independent.

s^2_e's are the prior constants.

Wong also proposed a prior of the following form:

$\underline{m}$ same as before.

For each column element of A, independently and hierarchically,

$\underline{a}_e | a^*_e \; : \; N_p(a^*_e \underline{1}_p \, , \, s^2_e I_p),$

$a^*_e \; : \; N(0 \, , \, s^2_{*e}), \; s^2_{*e} \longrightarrow$ infinite.

Empirical Bayes estimation of the prior constants, namely,

s^2_e, e=1,2,...,r, is discussed in his paper briefly.

4. Posterior distributions

L1

After integrating a^* and s^2 out,

$$f(A,D,C_F|S) = L_1(A,D)$$

$$\times \prod_* [\, [(a_{je} - a_{..})^2 + dw/n_A]^{-(nA+d-1)/2} \,]$$

$$\times |C_F|^{-(g-r-1)/2} \mathrm{Exp}[(-1/2)\mathrm{tr}(RC_F^{-1})]$$

$$\times \prod_{j=1}^{p} [\, d_j^{-(hj+2)/2} \,]$$

$$\times \mathrm{Exp}[(-1/2)\sum_{j=1}^{p} [\, h_j v_j / d_j] \,],$$

where $\prod_*$ denotes the product of those n_A elements of A

which are not fixed.

L2

$$f(A,D,C_F|S) = L_1(A,D)$$

$$\times \prod_e [\ [(a_{je}-a^*_{je})^2+d_{je}w_{je}]^{(dje+1)/2}\ ]$$

$\times$ (the last three factors of L1).

L3

$$f(A,D,C_F|S) = L_1(A,D)$$

$$\times\ |C_F|^{(r+1)/2}$$

$\times$ (the last two factors of L1)

L41, L42, and L43

$f(A,D,C_F|S)$ of these cases are the same as the ones in L1, L2, and L3, respectively, with the last two factors replaced by $\prod_{j=1}^{p}[Exp[(-1/2)v_j/d_j]\ ]$.

PR

$$f(A,D|S) = L_1(A,D)$$

$$\times\ |D|^{-h}Exp[(-1/2)(\underline{a}-\underline{a}^*)'G(\underline{a}-\underline{a}^*)].$$

KP

$$f(A,D|S) = L_1(A,D)f(A,\Omega),$$

where $f(A,\Omega)$ is defined in the above.

The marginal distribution of A is proportional to

$$f(A|S) \propto {}_1F_1 |AA'|^{-(h+N-1)/2}$$

$$\times \operatorname{Exp}[(-1/2)\operatorname{tr}(G(A-A_2)V_2(A-A_2)')],$$

where ${}_1F_1$ is the confluent hypergeometric function

of the arguments $(h+N)/2$, $h+N+p$, $-1/2(AA')^{-1}Q(A)+S_0$,

defined in Herz(1955).

WO

After marginalization of $\underline{m}$, the conditional

distribution of A given D and Y is proportional to

$$f(A|D,Y) \propto L_1(A,D)$$

$$\times \operatorname{Exp}[(-1/2)\sum_{e=1}^{r}[\ \underline{a}_e'\underline{a}_e/s_e^2]\ ].$$

5. Evaluation of the posterior distribution

Lee(1981) suggests the use of the joint mode as an estimate of each parameter. A numerical method analogous to the ML solution is proposed. However, this provides the location information only. Press(1982) does not suggest which values to be used as estimates. Kaufman and Press(1973) suggest the exact evaluation of the marginal mode and the normal approximation of the density. The apporoximation of the form

$$\underline{a}|S \ :\ N_{pr}(\ \underline{a}^+,\ S_1\ ) \text{ truncated so that}$$

$$(N/(N-1))S-AA' \text{ is p.d.},$$

where $\underline{a}$ is the column roll-out of A,

$S_1^{-1} = [V_1 \times S_{.}^{-1}] + [V_2 \times G]$,

$S_{.} = (N/(N-1)S/(h+N-1)$,

$\underline{a}^{+} = S_1[(V_1 \times S_{.}^{-1})\underline{a}_1 + (V_2 \times G)\underline{a}_2]$.

$\underline{a}_1$ and $\underline{a}_2$ are the column roll out of A_1

and A_2, respectively,

and x denotes the Kronecker product.

Wong(1980) does not follow the usual Bayesian approach. Instead, he suggests the maximization of

$L(D|Y) = \int f(A|D,Y)\, dA$,

which is the marginal likelihood of D. He suggests the use of the EM algorithm with numerical integration with respect to A above. For another set of the priors he also suggests the same algorithm to perform the hyperparameter estimation.

Discussion

As for the choice of the likelihood all the authors but Wong started with the sample dispersion matrix. This is due to the fact that $\underline{m}$ is of no interest in typical applications. However, as shown in the posterior distribution of Wong's method, the posterior distribution based on the likelihood of S, namely, L_1, can be obtained as a marginal posterior distribution of the L_2 based posterior distribution. In the limiting case when $\Omega_m^{-1} \longrightarrow$ zero, it can be

shown that with the normal prior of $\underline{m}$, the posterior marginal of A and D (and C_F) is the product of the Wishart kernel and the prior of A and D (, and C_F). Therefore, the method based on L_2 seems more general. On the contrary, the grand mean is so well estimated by the sample mean, it may be better to exclude $\underline{m}$ from the set of parameters of interest in order to make the resulting posterior distribution simple.

The assumption employed in Kaufman and Press(1973) that the off diagonal elements of D are not zero is a questionable one. It clearly contradicts the usual assumption of weak local independence, (2.1.11), or its strong form, (2.2.5). However, as Kaufmann and Press suggest, it accounts for the possibility of specification error, that is, if the posterior distribution of D is not concentrated on the diagonal matrix then it implies that the number of dimensions specified, r, is too small to explain the covariance structure of the form in (2.1.9). Therefore, as long as we can evaluate the posterior distribution of D, the nonzero off-diagonal assumption of D seems to provide useful information for determining the number of dimensions.

As for the form of prior distribution, the most interesting contrast exists between Lee's treatment of A and that of Press. That is, while Lee assumes those elements of A for which we have strong information to be fixed, Press places strong prior on those elements. Taking the specification error problem into account, Press' approach seems superior.

Finally, when the posterior distribution is complex, there always exists a problem of the choice between the exact modal estimate and the approximation of the posterior distribution by some evaluatable density such as the multivariate

normal distribution. Although the exact evaluation of the (joint) mode provides an exact value of one indicator of the location of the posterior distribution, it does not provide the dispersion information at all. On the other hand, if we approximate the posterior distribution we can have both location and dispersion information with less accuracy in the sense that those are the approximations. In face of the complexity of the posterior distribution this is an open question.

CHAPTER III

REVIEW OF THE CLASSICAL FIXED FACTOR ANALYTIC (FFA) MODEL

Model

As stated before the common factors $\underline{f}$ are treated as random latent variables in the RFA model. This is due to the fact that the RFA model was developed in close relation to the classical test theory model in which each subject is often treated as a random observation and the assumption of normality is to some extent reasonable. However, there are some areas in which it is impossible to have a random sample of subjects, yet the model (2.1.1) seems reasonable. Also, the value of $\underline{f}$ associated with each object, which is by definition impossible to estimate in the usual statistical sense under the RFA model, is often needed in many applications. In this chapter a model in which the common factors are treated as fixed quantities, namely, the Fixed Factor Analysis (FFA) model, will be reviewed.

The FFA model starts with (2.1.1) and its matrix equivalent form

(3.1.1) $\quad Y = \underline{1}_N\underline{m}' + FA' + U,$

or, equivalently,

$$\underline{y}_{(j)} = m_j\underline{1}_N + F\underline{a}_j + \underline{u}_{(j)}, \; j=1,2,\ldots,p,$$

where $Y = [\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_N]'$

$$= [\underline{y}_{(1)}, \underline{y}_{(2)}, \ldots, \underline{y}_{(p)}],$$

$$F = [\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_N]'$$

$$= [\underline{f}_{(1)}, \underline{f}_{(2)}, \ldots, \underline{f}_{(r)}],$$

$$A = [\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_p]',$$

$$\underline{m} = [m_1, m_2, \ldots, m_p]',$$

$$U = [\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_N]'$$

$$= [\underline{u}_{(1)}, \underline{u}_{(2)}, \ldots, \underline{u}_{(p)}].$$

(When we refer to the columns of each matrix we attach parentheses to the subscript, that is, $\underline{f}_{(e)}$ and $\underline{f}_i$ stand for the 1^{th} column and the i^{th} row of the F matrix, respectively.)

In order to avoid the redundancy of the parametrization the restrictions

(3.1.2) $\underline{1}'F = \underline{0}'$,

and

(3.1.3) $(1/N)F'F = C_F$, where $\mathrm{diag}(C_F) = I_r$,

are enforced. Since these restrictions are the counterparts of (2.1.3) and (2.1.4) in the RFA model, there still exists the rotational indeterminacy if C_F is identity. The distributional assumption

(3.1.4) $\underline{u}_{(j)}$: $N_N(\ \underline{0}\ ,\ d_j I_N\ )$, $j=1,2,\ldots,\ p$,

is usually made. The equivalent form of this assumption is

(3.1.5) $\underline{u}_i$: $N_p(\ \underline{0}\ ,\ D\ )$, $i=1,2,\ldots,N$, i.i.d.,

where D is the p x p matrix whose diagonal elements consist of d_j's.

From (3.1.1) and (3.1.4) it is clear that the FFA model is equivalent to

the multivariate regression model with <u>unknown</u> regressor matrix and with diagonal error dispersion matrix, or to the m-group regression model with <u>unknown common</u> regressor matrix with heteroschedastic error variances. [Novick, et., al. (1972).]

The likelihood of the set of parameters, namely, F, A, $\underline{m}$, and D, is given by

(3.1.6) $L_3(\underline{m},F,A,D|Y)$

$$= \prod_{j=1}^{p}[\ f(\underline{y}_{(j)}|m_j,F,\underline{a}_j,d_j)\ ]$$

$$= \prod_{j=1}^{p}[\ d_j^{-N/2}\mathrm{Exp}[(-1/2d_j)Q_j]\ ],$$

where

$$Q_j = (\underline{y}_{(j)}-m_j\underline{1}-F\underline{a}_j)'(\underline{y}_{(j)}-m_j\underline{1}-F\underline{a}_j),$$

or, equivalently,

(3.1.7) $L_3(\underline{m},F,A,D|Y) = |D|^{-N/2}\mathrm{Exp}[(-1/2)\mathrm{tr}[D^{-1}Q],$

where

(3.1.8) $Q = (Y-\underline{1}\underline{m}'-FA')'(Y-\underline{1}\underline{m}'-FA').$

As pointed out by Anderson and Rubin(1956), however, this likelihood is not bounded above, that is, if any one of the quadratic forms, Q_j, say, in the exponent is equal to zero the likelihood goes to infinity. The simplest way to avoid this problem is to extend the within-variable homoscedasticity assumption (3.1.4) to across-variable homoscedasticity, that is,

(3.1.9) $\underline{u}_{(j)}\ :\ N_N(\underline{0}\ ,\ dI_N\ ),\ j=1,2,\dots,p,$

where d is a scalar common to all the variables.

The MLE under this model with the restrictions (3.1.2) and (3.1.3) with $C_F = I_r$ is given by the Eckart-Young (1936) decomposition of the data matrix Y after column centering. That is, the estimates of $\underline{m}$, F, and A are given, respectively, by the column means of Y matrix, the first r eigen vectors of JYY'J normalized so as to satisfy (3.1.3), and the first r eigen vectors of the sample dispersion matrix S normalized so that A'A = L, the eigen value matrix. This is also the least squares solution for the model (3.1.1) and the resulting F matrix is equivalent to the first r principal component scores of the matrix JY. The distributional assumption (3.1.9), which assumes that all the variables have equal unique variances, however, seems to be too restrictive even if the original variables have the same variances. A better way is to assume either that the unique variances are known or that they are proportional to some known constants. For example, in case of supposedly unidimensioanl tests, the assumption that the error variance of each test is proportional to the length, as in Feldt (1975), may be used.

Another way to avoid the problem is to have replications. Denoting the k^{th} replication by $Y^{(k)}$ and the corresponding error term by $U^{(k)}$, k=1,2,...,q, the likelihood of the parameters given q replications is expressed as

(3.1.10) $$L_3(\underline{m},F,A,D|Y^{(1)},Y^{(2)},\ldots,Y^{(q)})$$

$$= |D|^{-qN/2}\mathrm{Exp}[(-1/2)\mathrm{tr}[D^{-1}Q_+]],$$

where

(3.1.11) $$Q_+ = \sum_{k=1}^{q}[\ Q^{(k)}\ ],$$

and

(3.1.12) $\quad Q^{(k)} = (Y^{(k)} - \underline{1}\underline{m}' - FA')'(Y^{(k)} - \underline{1}\underline{m}' - FA')$.

Since the diagonal elements of Q_+ cannot be zero (assuming each $Y^{(k)}$ is distinct) the likelihood is bounded above. The following estimating equations are given by taking the derivatives of the likelihood plus the Lagrange term to enforce (3.1.3) with $C_F = I_r$,

$$tr[(F'F - (N)I_N)L^*],$$

where L^* is the r x r symmetric unknown matrix,

and setting them equal to zero. (See the Appendix for matrix differentiation.)

(3.1.13) $\quad \underline{m} = (1/N)Y^{\cdot\prime}\underline{1}$,

where $Y^{\cdot} = (1/q) \sum_{k=1}^{q} [\, Y^{(k)} \,]$.

(3.1.14) $\quad D = (1/qN)\mathrm{diag}(Q_+)$.

(3.1.15) $\quad A = (1/N)Y^{\cdot\prime}JF$.

(3.1.16) $\quad F = JY^{\cdot}D^{-1}A(A'D^{-1}A + L^*)^{-1}$.

The last two equations reduce to

(3.1.17) $\quad RD^{-1/2}AT = D^{-1/2}ATL$,

where

$$R = (1/N)\, D^{-1/2}Y^{\cdot\prime}JY^{\cdot}D^{-1/2},$$

L is a r x r arbitrary diagonal matrix,

and T is a r x r arbitrary orthonormal matrix,

and

(3.1.18) $FT = N\, JY'D^{-1}AL^{-1}$.

Arbitrariness of T and L accounts for the orthogonal rotation. Therefore, by setting $T = I_r$, the conditional estimate of A given D is given as the eigen vectors of R normalized such that $A'D^{-1}A = L$, the eigen values.

Some other treatments of this problem have been proposed in Anderson and Rubin(1956), Anderson (1984), and McDonald(1979b). While the former two consider the estimation of the factor loadings and the error variances on the basis of the noncentral Wishart distribution with the restrictions (3.1.2) and (3.1.3) with $C_F = I_r$, the latter estimates the factor scores as well as the factor loadings and the error variances by maximizing the likelihood ratio. However, the method fails to produce unique estimates of the factor scores in the sense that all the estimates that are produced by the Guttman/Kestelman formula in the Appendix also maximize the likelihood ratio.

A Bayesian treatment of this model will be proposed in the next Chapter.

The Congeneric Test Model

Consider the situation in which an instructor is to grade his/her students on the basis of, say, three examinations such as two midterms and one final. This is usually done by calculating a certain composite score such as

A) mean of raw scores,

or

B) mean of standardized scores.

When we assume that there is only one common factor, that is, $r = 1$, the FFA model becomes, by denoting the first column of F and A by $\underline{f}$ and $\underline{a}$, respectively,

(3.2.1) $Y = \underline{1}\underline{m}' + \underline{f}\underline{a}' + U,$

or

$$y_{ij} = m_j + f_i a_j + u_{ij}, \quad i=1,2,\dots,N, \; j=1,2,\dots,p,$$

which states, as in the general case, that the observed score is the sum of non-random and random parts, namely, $m_j+f_i a_j$ and u_{ij}, respectively, with a particular structure being enforced on the non-random part. However, the way we decompose the observed score into two parts is not unique. As pointed out by Lord and Novick 1968, Chapters 2 and 24, we can have another decomposition, namely, the classical test theoretic decomposition,

(3.2.2) $y_{ij} = t_{ij} + e_{ij}, \; i=1,2,\dots,N, \; j=1,2,\dots,p,$

where t_{ij} is defined as a non-random part over the propensity distribution, i.e., replications. These two decompositions sometimes contradict each other since the factor analytic decomposition is usually made without any reference to the replications. That is, the unique score in the factor analytic decomposition may contain some of the non-random elements, namely, the specific score of each variable defined over replications. However, when we deal with only one set of observations, Y, it is impossible to distinguish the specific score from the unique score. Therefore, we proceed for the present as if the specific scores are zero.

With this understanding in mind (3.2.1) states that the non-random part of the observed score of each test, the true score, is linearly related to each

(4.2.17) $L_{D'} = \sum_{j=1}^{p}[\ L_{Dj'}\ ] + \text{const.},$

where

(4.2.18) $L_{Dj'} = -n \ln(s/2) + 2 \ln \text{Gamma}(n/2) + (n+2) \ln d_j + s/d_j.$

For the second stage prior distributions we only assume the mutual independence

(4.2.19) $f(\ H \mid S\ ) = f(\ H_F, H_A, H_D \mid S)$

$$= f(\ H_F \mid S_F\ )\ f(\ H_A \mid S_A\ )\ f(\ H_D \mid S_D\ ),$$

and do not elaborate the specific form until it becomes necessary. When a priori zeros are to be specified they should be so specified here. The easiest way is to assume a spike function as the second stage prior distribution of the hyperparameters whose elements are assumed to be concentrated around zero.

Posterior Joint Distribution

Using the Bayes' theorem the posterior joint distribution of F, A, D, and H is given as

(4.3.1) $f(\ F, A, D, H \mid S, Y\ ) = f(Y|F,A,D)\ f(F,A,D|H)\ f(H|S),$

Minus twice the log of this density is given by

(4.3.2) $L = L_{FAD} + L_F + L_A + L_{D'} + L_H + \text{const.}$

$$= \sum_{i=1}^{N}[\ L_{AFi} + L_{Fi}\ ] + L_A + L_D + L_H + \text{const},$$

$$= \sum_{j=1}^{p}[\ L_{FAj} + L_{Aj} + L_{Dj}\ ] + L_F + L_H + \text{const},$$

where

(4.3.3) $L_{FAD} = \text{tr}(\ Y-FA'\ )D^{-1}(\ Y-FA'\ )' + N \ln |D|,$

(4.2.11) $f(\ D \mid H_D\ ) = \prod_{j=1}^{p}[\ f(\ d_j \mid H_D\ )\ ]$,

where

(4.2.12) $f(\ d_j \mid H_D\ ) = (s/2)^{n/2}/\mathrm{Gamma}(n/2)\ d_j^{(-1/2)(n+2)}$

$$\times \mathrm{Exp}[\ (-1/2)s/d_j\ ],\ j=1,2,\ldots,p,$$

where Gamma(x) is the gamma function.

Note that $\underline{f}$ and $\underline{a}$, which do not have any subscript or which do have subscript k, are not parameters but hyperparameters.

The corresponding minus twice the log densities are,

(4.2.13) $L_F = \sum_{i=1}^{N}[\ L_{Fi}\ ] + \text{const.}$,

where

(4.2.14) $L_{Fi} = \ln |C_F| + \underline{f}_i{}'C_F^{-1}\underline{f}_i$,

if globally exchangeable, i=1,2,...., N,

$= \ln |C_{Fk}| + (\underline{f}_i-\underline{f}_k)'C_{Fk}^{-1}(\underline{f}_i-\underline{f}_k)$,

if locally exchangeable and i belongs to subgroup k, $k=1,2,\ldots,G_F$,

(4.2.15) $L_A = \sum_{j=1}^{p}[\ L_{Aj}\ ] + \text{const.}$,

where

(4.2.16) $L_{Aj} = \ln |C_A| + (\underline{a}_j-\underline{a})'C_A^{-1}(\underline{a}_j-\underline{a})$,

if globally exchangeable, j=1,2,...., p,

$= \ln |C_{Ak}| + (\underline{a}_j-\underline{a}_k)'C_{Ak}^{-1}(\underline{a}_j-\underline{a}_k)$,

if locally exchangeable and j belongs to subgroup k, $k=1,2,\ldots,G_A$,

on the first dimension, and the members of the second group have high loadings on the second dimension, we may specify that $\underline{a}_1$ and $\underline{a}_2$ are concentrated around $[a_{11}, 0]$ and $[0, a_{22}]$, respectively. This is a similar treatment of a priori zeros described in Rubin and Thayer (1982) but more general in the sense that this does not force the parameters to be zero, avoiding a potential specification error.

The density functions are, respectively,

(4.2.7) $f(\ F \mid H_F\ ) = \prod_{i=1}^{N} [\ f(\ \underline{f}_i \mid H_F\ )\ ]$

where

(4.2.8) $f(\ \underline{f}_i \mid H_F\ )$

$= |C_F|^{-1/2}\ \mathrm{Exp}[(-1/2)\underline{f}_i' C_F^{-1} \underline{f}_i]$,

if globally exchangeable, i=1,2,...., N,

$= |C_{Fk}|^{-1/2}\ \mathrm{Exp}[(-1/2)(\underline{f}_i - \underline{f}_k)' C_{Fk}^{-1} (\underline{f}_i - \underline{f}_k)]$,

if locally exchangeable and i belongs to subgroup k, k=1,2,....,G_F,

(4.2.9) $f(\ A \mid H_A\ ) = \prod_{j=1}^{p} [\ f(\ \underline{a}_j \mid H_A\ )\ ]$,

where

(4.2.10) $f(\ \underline{a}_j \mid H_A\ )$

$= |C_A|^{-1/2}\ \mathrm{Exp}[(-1/2)(\underline{a}_j - \underline{a})' C_A^{-1} (\underline{a}_j - \underline{a})]$,

if globally exchangeable, j=1,2,...., p,

$= |C_{Ak}|^{-1/2}\ \mathrm{Exp}[(-1/2)(\underline{a}_j - \underline{a}_k)' C_{Ak}^{-1} (\underline{a}_j - \underline{a}_k)]$,

if locally exchangeable and j belongs to subgroup k, k=1,2,....,G_A,

(4.2.6) $\underline{a}_j : N_r(\underline{a}_{k'}, C_{Ak'})$, iid, if j belongs to group k',

$k'=1,2,\ldots,G_A$,

where $\underline{a}_{k'}$ and $C_{Ak'}$ are, respectively, r x 1 and r x r.

In this case we have

$H_F = [(\underline{f}_k, C_{Fk}), k=1,2,\ldots,G_F]$,

and

$H_A = [(\underline{a}_k, C_{Ak}), k=1,2,\ldots,G_A]$.

A locally exchangeable prior distribution for the error variance is not considered here since in real application they can almost always be considered to be globally exchangeable. We denote the number of observations in the k^{th} locally exchangeable group by n_{Fk} and the number of variables in the k'^{th} locally exchangeable group by $n_{Ak'}$. It should be noted that (4.2.5) and (4.3.6) can be used independently. That is, for example, we may have a globally exchangeable prior on the factor scores and a locally exchangeable prior on the factor loadings. Also, by setting one of the inverse matrices of the dispersion hyperparameters equal to zero, we can handle those observations/variables whose a priori grouping information is not clear. That is, those observations/variables with zero precision hyperparameter matrix are treated as having uniform prior distributions.

When some of the location hyperparameters are assumed to be concentrated around zero we may specify this in the specification of the second stage prior distribution. For example, if there are two locally exchangeable groups of the variables and we believe that the members of the first group have high loadings

dispersion matrix,

(4.2.3) $\underline{a}_j : N_r(\underline{a}, C_A)$, $j=1,2,\ldots,p$, iid,

where $\underline{a}$ is the r x 1 vector of mean and C_A is the r x r

dispersion matrix,

(4.2.4) $d_j : X^{-2}(n, s)$,

where $X^{-2}(n,s)$ indicates the inverted chi square distribution with the degrees of freedom n and mean $s/(n-2)$.

With the previous notation we have

$H_F = [\underline{f}, C_F]$,

$H_A = [\underline{a}, C_A]$,

and

$H_D = [n, s]$.

The three sets of prior distributions, respectively, state that all the factor scores, factor loadings, and the error variances are globally exchangeable. Since the model is based on column centered data and considering the multiplicative redundancy between F and A, we set $\underline{f}=\underline{0}_r$ and $C_F=I_r$. The treatment of the oblique model will be stated later.

When the globally exchangeable prior distributions are not appropriate we may use the locally exchangeable prior distributions:

(4.2.5) $\underline{f}_i : N_r(\underline{f}_k, C_{Fk})$, iid, if i belongs to group k,

$k=1,2,\ldots,G_F$,

where $\underline{f}_k$ and C_{Fk} are, respectively, r x 1 and r x r,

variables require heavy knowledge of vocabulary and the rest are more content oriented we may a priori assume that there are two locally exchangeable groups of variables. Also, if we believe that there are some gender differences in terms of the factor scores but that the factor loadings are invariant across sex, we may assume the exchangeability of factor scores within boys and girls, but not globally. The model proposed here is general enough to handle any of these situation.

We consider the following forms for the prior distributions of the parameters.

(4.2.1) $f(F, A, D \mid H) = f(F|H_F)\, f(A|H_A)\, f(D|H_D)$,

where H_F, H_A, and H_D are the first stage hyperparameters of the prior distribution of F, A, and D, respectively. The independence assumption of F and (A, D) seems to be natural since knowledge of the characteristics of each subject usually does not affect knowledge of the characteristics of variables. The independence of A and D may not seem to represent the real situation considering the fact that the expected dispersion matrix of the observation is expressed as the sum of AA′ and D, but this is not the case since the prior distributions are to be specified prior to the data collection. That is, we argue that A and D are independent until we calculate the sample dispersion matrix.

For each component we first assume the following globally exchangeable prior distributions:

(4.2.2) $\underline{f}_i : N_r(\underline{f}, C_F)$, i=1,2,...,N, iid,

where $\underline{f}$ is the $r \times 1$ vector of mean and C_F is the $r \times r$

the formulae the range is not specified. The second stage hyperparameter S is typically specified to provide a relatively flat prior distribution for the H, expressing ignorance about the location. This type of prior distribution is called an exchangeable prior distibution. It should be noted that the first stage hyperparameters H can be constant if the distribution of H is really tight.

As a variation of the exchangeable prior presented above, we may have the following locally exchangeable prior.

P_k : $f(P_k \mid H_g)$, iid, if k belongs to a subset g,

H_g : $f(H_g \mid S)$, $g = 1,2,\ldots,G$, iid.

That is, we divide k parameters into G subsets and assume the exchangeability within each subset. Since the exchangeability assumption is crucial in real data analysis much caution should be excercized when it is incorporated.

In the factor analytic context there are N individuals, each of which is regarded as a population, and the location $\underline{f}_i$ of each individual is to be estimated. Also, each of the p variables represents p populations and their parameters $\underline{a}_j$'s and d_j's are to be estimated. Unless we are to perform some confirmatory study it is often difficult to specify the informative prior distributions for all of N+(r+1)p parameters. Also, it is often the case that we know that some of the tests or some of the subjects are very similar to each other prior to data collection. Therefore, the exchangeable prior seems to fit the typical application of the model very well. For example, if all the variables to be analyzed are supposed to measure reading skills we may a priori assume that those variables are exchangeable. However, if half of the

that is proper. Second, if we have very weak knowledge we use a noninformative prior distribution that is improper. The third case is that of an exchangeable prior distribution and is applicable when there are several of parameters which represent the same characteristics, say, location, of different populations and we believe those parameters are similar to each other. For example, when the means of m normal populations are to be estimated, we may know, a priori, that those m values are similar to each other but may not be able to precisely specify the 'mean' of these m values.

The relative similarity of prior values can be expressed as the following hierarchical forms. According to the model stated above it is assumed that

p_k : $f(p_k \mid H)$, $k=1,2,\ldots,m$, iid,

H : $f(H \mid S)$,

where $P = [p_1, p_2, \ldots, p_m]'$ is the m x 1 parameter vector of interest, H is the n1 x 1 first stage hyperparameter vector and S is the n2 x 1 second stage hyperparameter for the prior parameter H.

That is, we express the relative similarity of all parameters by assuming that they come from the same distribution and express the uncertainty of that distribution by the probabilistic structure of the hyperparameters. The prior distribution of the p_k's can be expressed without using H if we marginalize the joint distribution of P and H with respect to H. In this case the prior distribution of P can be written as

$$f(P \mid S) = \int f(P \mid H)\, f(H \mid S)\, dH.$$

Since it is usually the case that $n2 < n1 < m$, it should be easier to specify the prior constant S. It should be understood that the range of integrations is the domain of the variables to be integrated out. To avoid complexity in

$$f(\underline{y}_{(j)} \mid F, A, D) = f(\underline{y}_{(j)} \mid F, \underline{a}_j, d_j)$$

$$\propto d_j^{(-1/2)N}$$

$$\times \text{Exp}((-1/2)\, (\underline{y}_{(j)}-F\underline{a}_j)'(\underline{y}_{(j)}-F\underline{a}_j)/d_j).$$

Here, the symbol $\propto$ is again used to denote the proportionality.

Minus twice the log likelihood is

(4.1.5) $L_{FAD} = -2 \ln f(Y \mid F, A, D)$

$$= \text{tr}(Y-FA')D^{-1}(Y-FA')' + N \ln |D| + \text{const.}$$

$$= \sum_{i=1}^{N}[L_{AFi}] + N \ln |D| + \text{const.},$$

where

$$L_{AFi} = (\underline{y}_i-A\underline{f}_i)'D^{-1}(\underline{y}_i-A\underline{f}_i),$$

$$= \sum_{j=1}^{p}[L_{FAj}] + N \ln |D| + \text{const.},$$

where

$$L_{FAj} = (\underline{y}_{(j)}-F\underline{a}_j)'(\underline{y}_{(j)}-F\underline{a}_j)/d_j.$$

In the following sections the subscript i always refers to the observations, and the subscript j to the variables.

Prior Distributions

In general there are three ways to express our prior beliefs about the parameters of interest in a Bayesian estimation procedure. First, if we have strong knowledge of the parameters we use an informative prior distribution

where

$\underline{y}_{(j)}$ is the N x 1 vector of centered observations for the variable j,

$\underline{u}_{(j)}$ is the N x 1 vector or the error terms for the variable j,

$F = [\ \underline{f}_1, \underline{f}_2, \ldots, \underline{f}_N\ ]'$ is the N x r factor score matrix.

Collectively, we may write,

(4.1.3) $Y = F A' + U$,

where

$Y = [\ \underline{y}_{(1)}, \underline{y}_{(2)}, \ldots, \underline{y}_{(p)}\ ]$

$\quad = [\ \underline{y}_1, \underline{y}_2, \ldots, \underline{y}_N\ ]'$, N x p,

$U = [\ \underline{u}_{(1)}, \underline{u}_{(2)}, \ldots, \underline{u}_{(p)}\ ]$

$\quad = [\ \underline{u}_1, \underline{u}_2, \ldots, \underline{u}_N\ ]'$, N x p.

Therefore, the likelihood of F, A, and D given Y can be written as

(4.1.4) $f(\ Y \mid F, A, D\ ) =$

$$|D|^{(-1/2)N}\ \mathrm{Exp}(\ (-1/2)\ \mathrm{tr}[\ (Y-FA')D^{-1}(Y-FA')'\ ]\ )$$

$$= \prod_{i=1}^{N}[\ f(\ \underline{y}_i \mid F, A, D\ )\ ],$$

where

$$f(\ \underline{y}_i \mid F, A, D\ ) = f(\ \underline{y}_i \mid \underline{f}_i, A, D\ )$$

$$= |D|^{-1/2}$$

$$\times\ \mathrm{Exp}(\ (-1/2)\ (\underline{y}_i - A\underline{f}_i)'D^{-1}(\underline{y}_i - A\underline{f}_i)\ )\ ,$$

$$= \prod_{j=1}^{p}[\ f(\ \underline{y}_{(j)} \mid F, A, D\ )\ ],$$

where

$\underline{m} = (1/N)Y'\underline{1}_N$,

and do not treat it as a model parameter in order to have simpler form. Therefore, it should be understood that the observations are centered, i.e., each observed variable has a zero sample mean.

We state the model as follows.

(4.1.1) $\underline{y}_i = A\underline{f}_i + \underline{u}_i$, $i=1,2,\ldots,N$,

$\underline{u}_i : N_p(\ \underline{0}_p,\ D\ )$, $i=1,2,\ldots,N$, iid,

where

$\underline{y}_i$ is the p x 1 vector of the column centered observations for subject i,

$\underline{u}_i$ is the p x 1 vector of the error terms for subject i,

$\underline{f}_i = [\ f_{i1},\ f_{i2},\ \ldots,\ f_{ir}\ ]'$, $i=1,2,\ldots,N$,

is the r x 1 factor score vector for observation i,

$A = [\ \underline{a}_1,\ \underline{a}_2,\ \ldots,\ \underline{a}_p\ ]'$

where $\underline{a}_j = [\ a_{j1},\ a_{j2},\ \ldots,\ a_{jr}\ ]'$, $j=1,2,\ldots,p$,

is the r x 1 factor loading vector for the variable j,

is the p x r matrix of the factor loadings,

and

$D = \mathrm{diag}[\ d_1,\ d_2,\ \ldots,\ d_p\ ]$ is the p x p diagonal

dispersion matrix of the error term.

The diagonality of the D matrix enables us to reexpress the model variable wise as follows.

(4.1.2) $\underline{y}_{(j)} = F\underline{a}_j + u_{(j)}$, $j=1,2,\ldots,p$,

$\underline{u}_{(j)} : N_N(\ \underline{0}_N,\ d_j I_N\ )$, $j=1,2,\ldots,p$, iid,

CHAPTER IV

THE BAYESIAN FACTOR ANALYSIS

In this chapter a new Bayesian factor analytic model, where the factor scores as well as the factor loadings and the error variances are treated as one of the parameters to be estimated, is developed. The presentation is fully Bayesian in the sense that all the parameters have prior distributions and the inference is based on their posterior distributions. After the description of the model and the prior distributions the posterior joint distribution of all the parameters is derived. Then, the joint distribution is marginalized and/or conditionalized to obtain the point estimates. This enables us to reduce the factor score indeterminacy problem to the usual problem of the choice of point estimate, namely, mean, mode, or something else of the posterior distribution of the factor scores. Because the EM algorithm and its variatinos are used for the marginalization of the posterior joint distribution a brief description of the algorithm is provided after the derivation of the posterior distribution.

The Model

In this chapter a model similar to the one used in the fixed factor analysis is used. Since $\underline{m}$ is defined as the grand mean we replace it by the sample mean, namely,

section. Thus, within the framework of MLE it is impossible to estimate the parameters of the general congeneric model. Some modifications such as the across test homoscedasticity in (3.1.9) or

(3.2.7) $a_j = 1,\ j=1,2,\ldots,p,$

and d_j is proportional to the observed score variance of test j,

are necessary to deal with the form more general than the parallel test model. With the former assumption, with some restrictions concerning the scale and the origin of the true score such as (3.1.2) and (3.1.3), we have the first eigen vector of JYY'J as the MLE of $\underline{f}$, and with the latter, we have B).

Therefore, if we are to calculate a composite score based on the general congeneric test model, the Bayesian unidimensional FFA model proposed in the next chapter is the only possible way. It should be also noted that Lindley(1971a) has proposed a Bayesian solution to the parallel test model with an exchangeable prior distribution of the factor score.

other. This is sometimes referred to the congeneric test model, Kristof(1974), Feldt(1975), and contains the parallel test model or the tau-equivalent test model as a special case. That is, if all the a's are equal we have the essentially tau-equivalent test model, if, in addition, all the m's are equal we have the tau-equivalent test model, and if, in addition, all the unique variances are equal we have the parallel test model. Although classical test theory does not assume the strong form of the distributional assumption nor within test homoscedastisity such as

(3.2.3) $u_{ij} : N(0 , d_j)$, $i=1,2,\ldots,N$, $j=1,2,\ldots,p$,

we assume this to facilitate the comparison. This assumption, which specifies the characteristic of the error term associated with each test, does not seem to be strong compared to the normality assumption of each subject's true score used in the RFA model.

With assumption (3.2.3) and the parallel test assumption,

(3.2.4) $m_j = 0$, $j=1,2,\ldots,p$,

(3.2.5) $a_j = 1$, $j=1,2,\ldots,p$,

and

(3.2.6) $d_j = d$, $j=1,2,\ldots,p$,

it can be shown that the MLE of the factor score is given by A) in the example above.

However, if we allow the unique variance to vary freely in order to deal with the tau-equivalent models it can be shown that the MLE does not exist even with the restrictions (3.2.4) and (3.2.5). This is due to the same unboundedness of the L_3 likelihood function described in the previous

$$(4.3.4)\quad L_{AFi} = (\underline{y}_i - A\underline{f}_i)'D^{-1}(\underline{y}_i - A\underline{f}_i),$$

$$(4.3.5)\quad L_{FAj} = (\underline{y}_{(j)} - F\underline{a}_j)'(\underline{y}_{(j)} - F\underline{a}_j)/d_j,$$

$$(4.3.6)\quad L_D = \sum_{j=1}^{p}[\ L_{Dj}\ ],$$

$$(4.3.7)\quad L_{Dj} = L_{Dj'} + N \ln d_j$$

$$= -n \ln(s/2) + 2 \ln \mathrm{Gamma}(n/2) + (N+n+2) \ln d_j + s/d_j,$$

and

$$(4.3.8)\quad L_H = -2 \ln f(\ H \mid S\ ).$$

Unlike the L_3 likelihood in Chapter III the posterior joint distribution is not unbounded above due to the term s/d_j introduced by the prior distribution of the error variance. However, it is found that, unless we have the value of s which is comparable to the magnitude of the residual sum of squares,

$$RSS_j = (\underline{y}_{(j)} - F\underline{a}_j)'(\underline{y}_{(j)} - F\underline{a}_j),$$

the joint mode of (4.3.1) exists in the region where some of the error variances are close to zero.

Note that (4.3.1) also gives various conditional distributions by droping the factors which consist purely of the parameters/hyperparameters on which we would like to condition. In terms of (4.3.2), minus twice the log posterior conditional distribution of, for example, the factor scores and the factor loadings given the error variances and the hyperparameters is given by $L_{FAD}+L_F+L_A$+const.

Posterior Marginal Distributions

When the globally exchangeable prior is used either for the factor scores or for the factor loadings, following two marginal distributions can be derived analytically.

(4.4.1) $f(F, D, H \mid S, Y) = \int f(F, A, D, H \mid S, Y)\, f(A \mid S)\, dA$

$$= \int f(Y|F,A,D) f(A|H_A) dA\ f(F|H_F) f(D|H_D) f(H|S)$$

$$= \prod_{j=1}^{p} [\ |\ FC_A F' + d_j I_N\ |^{-1/2}\ x$$

$$\text{Exp}(\ (-1/2)(\underline{y}_{(j)} - F\underline{a})'(FC_A F' + d_j I_N)^{-1}(\underline{y}_{(j)} - F\underline{a})\ )\ ]$$

$$x\ f(F|H_F)\ f(D|H_D)\ f(H|S).$$

(4.4.2) $f(A, D, H \mid S, Y) = \int f(F, A, D, H \mid S, Y)\, f(F \mid S)\, dF$

$$= \int f(Y|F,A,D) f(F|H_F) dF\ f(A|H_A) f(D|H_D) f(H|S)$$

$$= |\ AC_F A' + D\ |^{(-1/2)N}\ \text{Exp}(\ (-1/2)\ tr(\ Y'Y D^{-1}\ )\ )$$

$$x\ f(A|H_A)\ f(D|H_D)\ f(H|S).$$

Also, with the globally exchangeable prior of the error variances, we have

(4.4.3) $f(F, A, H \mid S, Y) = \int f(F, A, D, H \mid S, Y)\, dD$

$$= \prod_{j=1}^{p} [\ (RSS_j + s)^{-(N+n)/2}\ ]$$

$$x\ p\ (s/2)^{n/2}\ \text{Gamma}((N+n)/2)\ /\ \text{Gamma}(n/2)$$

$$x\ f(F|H_F)\ f(A|H_A)\ f(H|S),$$

where $RSS_j = (\underline{y}_{(j)} - F\underline{a}_j)'(\underline{y}_{(j)} - F\underline{a}_j)$,

and Gamma(x) is the Gamma function.

Note that the posterior marginal distribution of A, D, and H is

essentially equivalent to the familiar likelihood on which the usual MLE is based. In this sense our approach includes both the random and the fixed factor analytic model reviewed in the previous chapters.

Theoretically, when we are interested in the estimation of the factor scores and the factor loadings we may be able to use the joint mode of the marginal distribution of the factor scores and the factor loadings as the point estimates. However, since the distribution given in (4.3.3) has the same tendency as the joint posterior distribution given in (4.3.1), it is not desirable to use the mode of (4.3.3) as the estimate. That is, the mode is close to the region where some of the d_j's are zero. (It can be shown that (4.4.1) also has the same characteristic.) On the other hand, the marginal distributions given in (4.4.1) and (4.4.2), contain the location parameter F or A and the scale parameter D together. Therefore, the modes of these densities are, in this sense, still joint modes (e.g. (F,D) or (A,D)) and suffer the criticism of O'Hagan (1976), or Fienberg (1972). Also, Rubin and Thayer (1982), who derived the MLE from the equivalent likelihood, note that 'estimation of variances should be from their marginal likelihood.' The point is that the joint mode or the joint MLE tends to underestimate the error variances due to the lack of degrees of freedom adjustment, often resulting in the Heywood case. Therefore, some additional marginalization is necessary.

Since we have the hyperparameters in the posterior distribution we first marginalize all the parameters in order to have point estimates of the hyperparameters. That is, the point estimates of the hyperparameters are given by the mode of

$$(4.4.4) \quad f(H \mid S, Y) = \int\int\int f(F, A, D, H \mid S, Y) \, dF \, dA \, dD.$$

Then, conditioned on the obtained values of the hyperparameters, the point estimate of the error variance is derived as the mode of

$$(4.4.5)\quad f(\ D \mid H,\ S,\ Y\ ) = \int\int f(\ F,\ A,\ D \mid H,\ S,\ Y\ )\ dF\ dA.$$

Then, conditioned on the point estimates of the error variances and the hyperparameters, we estimate the factor scores and the factor loadings as the mode of

$$(4.4.6)\quad f(\ F,\ A \mid D,\ H,\ S,\ Y\ ),$$

$$(4.4.7)\quad f(\ F \mid D,\ H,\ S,\ Y\ ) = \int f(\ F,\ A \mid D,\ H,\ S,\ Y\ )\ dA,$$

or

$$(4.4.8)\quad f(\ A \mid D,\ H,\ S,\ Y\ ) = \int f(\ F,\ A \mid D,\ H,\ S,\ Y\ )\ dF.$$

As O'Hagan (1976) noted, the above conditional modes are considered to be better estimate than the marginal mode. When the main interest of the analysis is the estimation of the factor loadings, the mode of (4.4.8) should be used. When the main interest of the analysis is the estimation of the factor scores the mode of (4.4.7) should be used. Finally, if both are of interest the mode of (4.4.6) should be used.

Since analytic marginalization is impossible we use some variations of the EM algorithm, Dempster, et. al. (1977), in the later section. A brief descriptions of the EM algorithm and its variations are presented in the next section.

The EM Algorithm and Its Variations

Given a set of random variables ($\underline{u}$, $\underline{v}$), and their joint density function, f($\underline{u},\underline{v} \mid \underline{p}$), where, $\underline{p}$ is the parameter which determines the density,

the EM algorithm can be used to find the maximum likelihood estimate (MLE) of $\underline{p}$ on the basis of the marginalized likelihood

$$g(\, \underline{u} \mid \underline{p} \,) = \int f(\, \underline{u}, \underline{v} \mid \underline{p} \,)\, d\, \underline{v}.$$

In the original article by Dempster, et.al,(1977), a set ($\underline{u}$, $\underline{v}$) is refered to complete data, $\underline{u}$, incomplete data, and $\underline{v}$, missing data. The terminology makes sense when $\underline{v}$ represents the portion of observations which are missing but in more general applications it should be understood that the random variables to be integrated out are refered to as missing data. (See the following examples.) The algorithm is an iterative process consisting of two steps, namely, the E-step and the M-step. Assuming that an estimate of $\underline{p}$, say, $\underline{p}_0$, is given, for each iteration the E-step calculates the conditional expectation of ln(f($\underline{u}$,$\underline{v}$ | $\underline{p}$)) given $\underline{v}$ and $\underline{p}_0$, namely,

$$E_V(\underline{u}|\underline{p}) = \int \ln(f(\underline{u},\underline{v}|\underline{p}))f(\underline{v}|\underline{u},\underline{p}_0)\, d\underline{v},$$

and in the M-step $E_V(\underline{u}|\underline{p})$ is maximized with respect to $\underline{p}$ assuming the parameters of the conditional distribution of $\underline{v}$, say $\underline{v}^*$, are constant. That is, although the parameter $\underline{v}^*$, which is a function of $\underline{u}$, and $\underline{p}_0$, is included in $E_V(\underline{u}|\underline{p})$ it is treated as a constant when the derivative of E_V is taken. Notation such as $E_V(\underline{u}|\underline{p})$ should not be confused with the simple expectation sign $\underline{E}[.]$. The successive application of these two steps will usually result in the MLE of the marginal likelihood $g(\underline{u}|\underline{p})$ when the process converges. It is assumed that some initial value of p is given prior to the

iteration. See Wu (1983) for the conditions required for the convergence.

When the complete data has the distribution which belongs to the regular exponential family the conditional expectation of the sufficient statistics, $\underline{t}(\underline{u},\underline{v})$, may be used for the calculation since the log likelihood can be writen as

$$\ln f(\underline{u},\underline{v}|\underline{p}) = \underline{p}'\underline{t}(\underline{u},\underline{v}) + \ln a(\underline{p}) + \text{const.},$$

where $a(\underline{p})$ is a function of $\underline{p}$ only,

and const. is a constant term which does not include $\underline{p}$.

When $\underline{p}$ has a prior distribution, $f(\underline{p}|H)$, the EM algorithm can be used to find the posterior mode of

$$g(\ \underline{p}\ |\ \underline{u},\ H\ ) = \int f(\ \underline{p},\ \underline{v}\ |\ \underline{u},\ H\ )\ d\ \underline{v}$$

$$= \int f(\underline{u},\underline{v}|\underline{p})d\underline{v}\ f(\underline{p}|H).$$

In this case, the E-step calculates the conditional expectation with respect to $f(\underline{v}|\underline{u},\underline{p},H)$, that is,

$$E_V(\underline{p}|\underline{u},H) = \int \ln[f(\underline{u},\underline{v}|\underline{p})]f(\underline{v}|\underline{u},\underline{p},H)d\underline{v},$$

and the M-step maximizes this with respect to $\underline{p}$ treating the parameter of the conditional distribution of $\underline{v}$ as a constant. When the process converges we presume to have the posterior marginal mode of $\underline{p}$ with $\underline{v}$ integrated out.

Now, suppose we have an observation Y, whose distribution is described by $f(Y|P)$, and a hierarchical prior distribution of P of the form $f(P|H)f(H|S)$, where H is the first stage hyperparameter, and S, the second stage hyperparameter. The posterior joint distribution of P and H is given by

$f(\,P,\,H \mid S\,) = f(Y|P)\, f(P|H)\, f(H|S),$

and, the posterior marginal distribution of H with P integrated out by

$g(\,H \mid S\,) = \int f(P,H|S)\, dP.$

In order to estimate the modal value of H of the marginal posterior distribution g, we can also use the EM algorithm by identifying

$\underline{u} = Y$, $\underline{v} = P$, and $\underline{p} = H$.

When the minimization of minus (twice) the log likelihood is preferred to the maximization due to its simplicity, its expectation,

$E_V(\underline{u}|\underline{p}) = (-2) \int \ln f(\underline{u},\underline{v}|\underline{p}) f(\underline{v}|\underline{u},\underline{p}) d\underline{v}$, should be calculated in the E-step and, in the M-step, it should be minimized with respect to $\underline{p}$.

One of the advantages of the EM algorithm over the direct maximization of $g(\underline{u}|\underline{p})$ is its simplicity. In the usual application, where $(\underline{u},\underline{v})$ have multivariate normal distribution, E_V is much easier to work with. That is, the maximization can be done analytically. Another advantage is its flexibility. Bock and Aitkin (1981) used its variation in order to have marginal MLE of the item parameters under the latent trait model where the conditional expectation of $\underline{v}$ is plugged into the complete log likelihood instead of taking its expectation. We propose another variation in order to marginalize with respect to several sets of missing data, say, $\underline{v}_1$, $\underline{v}_2$, ... in the later section.

In the following examples the actual application of the EM algorithm will be explained. The first example is a straightforward one where some of the observations are missing. The second example shows how we treat a subset of the parameters as the missing observations.

Example 1: Estimation of Multivariate Normal Parameters

Let Z be the n x p data matrix containing n observations from $N_p(\underline{m}, W)$. Now, assume a part of Z is missing. That is, let

$$Z = \begin{bmatrix} \underline{x}_1' \\ \underline{x}_2' \\ : \; : \\ : \; : \\ \underline{x}_{n1}' \\ \underline{y}_1' \\ \underline{y}_2' \\ : \; : \\ : \; : \\ \underline{y}_{n2}' \end{bmatrix},$$

where n=n1+n2, and

$\underline{x}_i$, i=1,2,...,n_1, be p x 1 vector of observation,

$\underline{y}_i$, i=1,2,...,n_2, be p x 1 vector where

$\underline{y}'_i = [\ \underline{a}_i' \ , \ \underline{b}_i' \ ]$,

$\underline{a}_i$, i=1,2,...,n2, and $\underline{b}_i$, i=1,2,...,n2, are,

respectively, p1 x 1 and p2 x 1 vectors and p=p1+p2.

It should be understood that $\underline{b}_i$'s are missing.

With the notation used in the previous section, we have,

$\underline{u}$ = [$\underline{x}_i$, i=1,2,...,n1, and $\underline{a}_i$, i=1,2,...,n2],

$\underline{v}$ = [$\underline{b}_i$, i=1,2,...,n2],

and

$\underline{p}$ = [$\underline{m}$ and W],

and we would like to estimate $\underline{m}$ and W on the basis of

the marginal likelihood

$$g(X,\underline{a}_i,\ i=1,2,\dots,n2 \mid \underline{m},W) = \int f(Z \mid \underline{m},W)\ d\ \underline{b}.$$

See Orchard and Woodbury (1972) for theoretical basis of this method.

Now, minus twice the log likelihood of ($\underline{m}$,W) given Z, the complete data, is

$$L = n \ln|W| + \sum_{i=1}^{N}((\underline{z}_i-\underline{m})'W^{-1}(\underline{z}_i-\underline{m})) + \text{const.}$$

The E-step.

1. Conditional Distribution of $\underline{b}_i$'s given X and A.

The conditional distribution of $\underline{b}_i$ given X and $\underline{a}_i$ is

$$\underline{b}_i \mid X, A \quad : \quad N_{p2}(\underline{b}_i^*, V^*),$$

where

$$\underline{b}_i^* = \underline{m}_2 + W_{21}W_{11}^{-1}(\underline{a}_i-\underline{m}_1)$$

$$V^* = W_{22}-W_{21}W_{11}^{-1}W_{12},$$

where

$\underline{m}$ = [$\underline{m}_1$', $\underline{m}_2$']',

1 x p1 1 x p2

and

$$W = \begin{bmatrix} W_{11}, & W_{12} \\ W_{21}, & W_{22} \end{bmatrix}.$$

2. Conditional Expectation of -2 log likelihood.

With obvious notation

$$W^{-1} = \begin{bmatrix} W^{11}, & W^{12} \\ W^{21}, & W^{22} \end{bmatrix},$$

the i^{th} term in the summation which includes $\underline{b}_i$ can be decomposed as

$$Q_i = (\underline{Y}_i - \underline{m})' W^{-1} (\underline{Y}_i - \underline{m})$$

$$= (\underline{a}_i - \underline{m}_1)' W^{11} (\underline{a}_i - \underline{m}_1)$$

$$+ 2 (\underline{a}_i - \underline{m}_1)' W^{21} (\underline{b}_i - \underline{m}_2)$$

$$+ (\underline{b}_i - \underline{m}_2)' W^{22} (\underline{b}_i - \underline{m}_2).$$

Therefore, the conditional expectation of Q_i is

$$\underline{E}(Q_i \mid X, \underline{a}_i, i=1,2,\ldots,n2)$$

$$= \underline{E}(Q_i \mid \underline{a}_i)$$

$$= (\underline{a}_i - \underline{m}_1)' W^{11} (\underline{a}_i - \underline{m}_1)$$

$$+ 2 (\underline{a}_i - \underline{m}_1)' W^{21} (\underline{b}_i^* - \underline{m}_2)$$

$$+ (\underline{b}_i^* - \underline{m}_2)' W^{22} (\underline{b}_i^* - \underline{m}_2)$$

$+ \mathrm{tr}W^{22}V^{*},$

$= (\underline{y}_i^{*}-\underline{m})'W^{-1}(\underline{y}_i^{*}-\underline{m}) + \mathrm{tr}W^{22}V^{*},$

where $\underline{y}_i^{*} = [\ \underline{a}_i',\ \underline{b}_i^{*}{}'\ ]'$.

(See Appendix for the expectation of quadratic form.)

Therefore,

$\underline{E}(L \mid X,\ \underline{a}_i, i=1,2,\ldots,n2) = E_B(\underline{m},W \mid X,A)$

$= \sum_{i=1}^{N}((\underline{x}_i-\underline{m})'W^{-1}(\underline{x}_i-\underline{m}))$

$+ \sum_{i=1}^{N}((\underline{y}_i^{*}-\underline{m})'W^{-1}(\underline{y}_i^{*}-\underline{m}))$

$+ n2\ \mathrm{tr}W^{22}V^{*} + \mathrm{const.}$

The M-step

By differentiating $E_B(\underline{m},W \mid X,\ \underline{a}_i, i=1,2,\ldots,n2)$

with respect to $\underline{m}$ and W^{ij}, ij= 11, 12, and 22,

(see the Appendix for the Matrix Differentiation,)

and solving the resulting normal equations, we have

$\underline{m}^{+} = (1/n)[\ X',\ Y^{*}{}'\ ]\ \underline{1}_n,$

$W_{ij}^{+} = (1/n)\ C_{ij},$ ij = 11, 12, and 21,

and

$W_{22}^{+} = (1/n)\ (\ C_{22} + n2\ V^{*}\ ),$

where C is the sample mean corrected sum of squares and cross product

matrix and C_{ij}'s are its partition similar to the one for W.

Note that $\underline{b}_i^*$'s and V^* are treated as constants when taking the derivatives. The successive application of these two steps will results in the marginal MLE of $\underline{m}$ and W with the missing data integrated out, namely, $\underline{m}^+$ and W^+, at the convergence point.

Example 2: Two-Way Random Effect ANOVA without Interaction

Let us consider the two way random effect ANOVA model without interactions.

$y_{ij} = a_i + b_j + e_{ij}$, i=1,2,...,p, j=1,2,...,q,

where

e_{ij} : N(0, s_E), i=1,2,...,p, j=1,2,...,q, iid,

a_i : N(m, s_A), i=1,2,...,p, iid,

and

b_j : N(0, s_B), j=1,2,...,q, iid.

Note that the overall mean is taken care of by the mean of parameters, a_i, namely, m. Also, to avoid the complexity, no replication is made. Writing the model using the linear regression form, we have

$\underline{y} = X\underline{\beta} + \underline{e}$,

where

$\underline{y}$ = [y_{11}, y_{12}, ..., y_{1q}, y_{21}, ..., y_{pq}]', pq x 1,

$\underline{e}$ = [e_{11}, e_{12}, ..., e_{1q}, e_{21}, ..., e_{pq}]', pq x 1,

$$X = \begin{bmatrix} \underline{1}_q & & & & I_q \\ & \underline{1}_q & & & I_q \\ & & \vdots & \vdots & \\ & & \vdots & \vdots & \\ & & & \underline{1}_q & I_q \end{bmatrix}, \quad pq \times (p+q),$$

$$\underline{\beta} = [\ \underline{a}',\ \underline{b}'\ ]', \quad (p+q) \times 1,$$

$$\underline{a} = [\ a_1,\ a_2,\ \ldots,\ a_p\ ]', \quad p \times 1,$$

$$\underline{b} = [\ b_1,\ b_2,\ \ldots,\ b_q\ ]', \quad q \times 1,$$

$$\underline{\beta} : N_{p+q}(\ \underline{m}\ ,\ C\ ),$$

$$\underline{m} = [\ m\underline{1}_p',\ \underline{0}_q'\ ]', \quad (p+q) \times 1,$$

$$C = \begin{bmatrix} s_A I_p & \\ & s_B I_q \end{bmatrix}, \quad (p+q) \times (p+q).$$

By integrating $\underline{\beta}$ out analytically, we see that $\underline{y}$ has the pq-variate normal distribution with

mean of $y_{ij} = m$,

and

$$\text{cov}(\ y_{ij}, y_{kl}\ ) = d_{ik}s_A + d_{jl}s_B + d_{ik}d_{jl}s_E,$$

where d_{ij} is the Kronecker's delta notation.

Instead of working directly on the marginal likelihood implied above, we proceed as follows.

Let us assume that [$\underline{y}$, $\underline{\beta}$] is the complete data and $\underline{\beta}$ is the missing

part. That is

$$\underline{u} = \underline{y},\ \underline{v} = \underline{\beta},\ \text{and}\ \underline{p} = [\ s_A,\ s_B,\ s_E,\ \text{and}\ \underline{m}\ ].$$

Minus twice the log likelihood of the complete data is

$$L = pq\ \ln(s_E) + (\ \underline{y}-X\underline{\beta}\ )'(\ \underline{y}-X\underline{\beta}\ )\ /\ s_E$$

$$+ \ln |C| + (\ \underline{\beta}-\underline{m}\ )'C^{-1}(\ \underline{\beta}-\underline{m}\ ) + \text{const.}$$

The E-step

1. Conditional Distribution of $\underline{\beta}$ given $\underline{y}$.

By rearranging two quadratic forms we see that

$$\underline{\beta} \mid \underline{y} : N_{p+q}(\ \underline{\beta}^*,\ V^*\ ),$$

where

$$V^* = (\ s_E^{-1}X'X + C^{-1}\ )^{-1},$$

and

$$\underline{\beta}^* = V^*\ (\ s_E^{-1}X'\underline{y} + C^{-1}\underline{m}\ ),$$

$$= [\ \underline{a}^{*\prime},\ \underline{b}^{*\prime}\ ],\ \text{say.}$$

(See the completion of sum of squares trick in the Appendix.)

2. Conditional Expectation of L given $\underline{y}$.

$$\underline{E}(\ L \mid \underline{y}\ ) = E_\beta(\ m,\ s_A, s_B, s_E \mid \underline{y}\ )$$

$$= pq\ \ln(s_E) + (\underline{y}-X\underline{\beta}^*)'(\underline{y}-X\underline{\beta}^*)/s_E + trX'XV^*/s_E$$

$$+ \ln |C| + (\underline{\beta}^*-\underline{m})'C^{-1}(\underline{\beta}^*-\underline{m}) + trC^{-1}V^* + \text{const}$$

$$= pq\ \ln(s_E) + (\ (\underline{y}-X\underline{\beta}^*)'(\underline{y}-X\underline{\beta}^*)+trX'XV^*\ )\ /\ s_E$$

and

$$\underline{a}_j^{+} = (F^{*\prime}F^{*})^{-1}F^{*\prime}\underline{Y}_{(j)}.$$

Therefore, conditioned on $\underline{Y}_{(j)}$, d_j, H, and S, we have

$$(4.6.3.3.)\quad \underline{a}_j \mid D,\ H,\ S,\ Y \quad : \quad N_r(\ \underline{a}_j^{*},\ Q_j^{*}\ ),$$

and these conditional distributions are independent for j=1,2,...,p. The joint density is denoted by $f^{*}(\ A \mid D,\ H,\ S,\ Y\ )$, which is the product of each normal density in (4.6.3.3). When some of the elements of D is zero the calculation of $Q^{*}_{\ j}$'s should be performed using the matrix inversion formula 2 in the Appendix.

As we noted, depending on the method of quasi marginalization, we can either take the expectation of $E_F(A,D,H|S,Y)$ with respect to the conditional distribution of A or we can analytically integrate A out. First we consider taking the expectation of E_F.

$$(4.6.3.4)\quad E_{AF}(D,H|S,Y) = \int E_F(A,D,H|S,Y)\ f^{*}(A|D,H,S,Y)\ d\ A$$

$$= \sum_{j=1}^{p}[\ \int\ (L_{FAj*}+L_{Aj}+N\underline{a}_j{}'V^{*}\underline{a}_j)\ d\ \underline{a}_j\ ]$$

$$+\ L_{F*} + L_D + L_H + \text{const},$$

$$= \sum_{j=1}^{p}[\ (\ RSS_j^{**}+tr[(F^{*\prime}F^{*}+N\ V^{*})Q_j^{*}]$$

$$+\ N\ \underline{a}_j^{*\prime}V^{*}\underline{a}_j^{*}\ )/d_j\ ]$$

$$+\ L_{A*} + L_{F*} + L_D + L_H + \text{const.},$$

where

Now consider $E_F(A, D, H \mid S, Y)$ in (4.6.2.1) as minus twice the log posterior density of A, D, and H given S and Y and denote the density by $f^*(A,D,H|S,Y)$. Then, by writing variable wise, we have

(4.6.3.1) $E_F(A, D, H \mid S, Y)$

$$= \sum_{j=1}^{p} [(RSS_j^* + N\underline{a}_j{}'V^*\underline{a}_j)/d_j + L_{Aj}]$$

$$+ L_{F*} + L_D + L_H + \text{const},$$

where

$$RSS_j^* = (\underline{y}_{(j)} - F^*\underline{a}_j)'(\underline{y}_{(j)} - F^*\underline{a}_j).$$

Again using the projection operator trick 1 and the sum of squares trick in the Appendix twice, we have

(4.6.3.2) $L_{FAj*} + L_{Aj} + N\,\underline{a}_j{}'V^*\underline{a}_j/d_j$

$$= (\underline{a}_j - \underline{a}_j^*)'Q_j^{*-1}(\underline{a}_j - \underline{a}_j^*)$$

$$+ \underline{a}_j^+{}'F^*{}'F^*\underline{a}_j^+/d_j + \underline{a}'C_A^{-1}\underline{a}$$

$$- \underline{a}_j^*{}'Q_j^{*-1}\underline{a}_j^* + \underline{y}_{(j)}{}'P\underline{y}_{(j)}/d_j,$$

where

$$L_{FAj*} = RSS_j^* / d_j,$$

$$Q_j^* = ((1/d_j)F^*{}'F^* + C_A^{-1} + (N/d_j)V^*)^{-1},$$

$$\underline{a}_j^* = Q_j^*((1/d_j)F^*{}'\underline{y}_{(j)} + C_A^{-1}\underline{a}),$$

$$P = I_N - F^*(F^*{}'F^*)^{-1}F^*{}',$$

where

$$L_{Fk*} = \sum_{i*k}[$$

$$(\underline{f}_i-\underline{f}_k^*)'C_{Fk}^{-1}(\underline{f}_i-\underline{f}_k^*)\ ]$$

$$+\ n_{Fk}\mathrm{tr}\ C_{Fk}^{-1}V_k^* \quad +\ n_{Fk}\ln|C_{Fk}|.$$

The sign $\sum_{i*k}$ denotes the sum over those i's which belong to subgroup k.

In this case, the quantities to be stored are

$$(4.6.2.6)\ Y'_kF_k^* = (\underline{y}_{+k}\underline{f}_k'C_{Fk}^{-1}+Y_k'Y_kD^{-1}A)V_k^*,$$

and

$$(4.6.2.7)\ F_k^{*\prime}F_k^* = V_k^*[\ A'D^{-1}\ Y_k'Y_k\ D^{-1}A$$

$$+\ A'D^{-1}\underline{y}_{+k}\underline{f}_k'C_{Fk}^{-1} + C_{Fk}^{-1}\underline{f}_k\underline{y}'_{+k}D^{-1}A$$

$$+\ n_{Fk}C_{Fk}^{-1}\underline{f}_k\underline{f}_k'C_{Fk}^{-1}\ ]V_k^*,$$

where

$$\underline{y}_{+k} = \sum_{i*k}[\ \underline{y}_i\ ].$$

Therefore, $\underline{y}_{+k}$ and $Y_k'Y_k$, $k=1,2,\ldots,G_F$ are sufficient in order to estimate the factor score weight.

Also, all the expressions in the following sections which contain the term $N\ V^*$ must be corrected in the similar manner as (4.6.2.5) as well as L_{F*}.

Quasi Marginalization with respect to the Factor Loadings

inversion formula 1 in the Appendix. (See Technical Notes section below.) As noted before in the example of the EM algorithm, the MLE by the EM algorithm does not produce negative estimates of the error variances since the V^* matrix is positive semi definite.

Also, since F^* always appears as $Y'F^*$ and $F^{*\prime}F^*$ it is not necessary to store all the factor score estimates. That is,

(4.6.2.2) $Y'F^* = Y'Y\, D^{-1}\, A\, V^*$,

and

(4.6.2.3) $F^{*\prime}F^* = V^*\, A'\, D^{-1}\, Y'Y\, D^{-1}\, A\, V^*$,

are to be stored in the calculation. With those quantities, RSS_j^* in (4.6.3.1) can be reexpressed as

(4.6.2.4) $RSS_j^* = [Y'Y]_{jj} - 2[Y'F^*]_j{}'\underline{a}_j + \underline{a}_j{}'[F^{*\prime}F^*]\underline{a}_j$.

In this sense the mean corrected sum of squares and cross product matrix Y'Y is sufficient for the estimation. (It is not sufficient for the estimation of the factor score, but sufficient for the scoring weight $W = D^{-1}AV^*$.)

When the locally exchangeable prior is used for the factor scores (4.6.2.1) should be replaced by

(4.6.2.5) $E_F(\ A,\ D,\ H \mid S,\ Y\ )$

$$= L_{FAD*} + tr(\ A'D^{-1}A \sum_{k=1}^{GF}[\ n_{Fk}V_k^*\ ]\ )$$

$$+ L_{F*} + L_A + L_D + L_H + \text{const.},$$

where

$$L_{F*} = \sum_{k=1}^{GF}[\ L_{Fk*}\ ],$$

(4.6.1.5) $F^*_k = (\underline{1}_{nFk}\underline{f}_k{}' C^{-1}_{Fk} + Y_k D^{-1} A) V^*_k$.

When some of the elements of D are zero the calculation of F^* should be performed using Lawley's trick in the Appendix.

Conditional Expectation of Minus Twice the Log Posterior Density

From (4.6.1.2) and (4.3.2) we have

(4.6.2.1) $E_F(A, D, H \mid, S, Y) = \int L f(F \mid A, D, H, S, Y) \, d F$

$= \sum_{i=1}^N [L_{AFi*} + L_{Fi*}] + N \operatorname{tr} A'D^{-1}AV^*$

$+ N \operatorname{tr} C_F^{-1}V^* + N \ln|C_F| + L_A + L_D + L_H + \text{const},$

where

$L_{AFi*} = (\underline{y}_i - A\underline{f}^*_i)'D^{-1}(\underline{y}_i - A\underline{f}^*_i),$

and

$L_{Fi*} = \underline{f}^{*\prime}_i C_F^{-1} \underline{f}^*_i,$

$= L_{FAD*} + N \operatorname{tr}[A'D^{-1}AV^*] + L_{F*} + L_A + L_D + L_H + \text{const},$

where

$L_{FAD*} = \operatorname{tr}(Y - F^*A') D^{-1} (Y - F^*A')',$

$L_{F*} = \sum_{i=1}^N [L_{Fi*}] + N \operatorname{tr} C_F^{-1}V^* + N \ln|C_F|$

Note that if we minimize this expectation with respect to A and D with uniform prior of A and D, we have the same estimate as given in Rubin and Thayer (1982). This can be shown by using Lawley's trick and the matrix

(4.6.1.2) $\underline{f}_i \mid \underline{y}_i$, A, D, H, S : $N_r(\underline{f}_i^*, V^*)$,

and these conditional distributions are independent for i=1,2,...,N.
The joint density, which is the product of the normal density given above is denoted by f(F | A, D, H , S, Y).

We also write

(4.6.1.3) $F^* = Y W$,

where $W = D^{-1}AV^*$.

When the locally exchangeable prior is used for the factor scores, (4.6.1.2) and (4.6.1.3) should be replaced by

(4.6.1.4) $\underline{f}_i^* = V_k^* (A'D^{-1}\underline{y}_i + C_{Fk}^{-1}\underline{f}_k)$,

if the observation i belongs to the k^{th} group,

where

$$V_k^* = (A'D^{-1}A + C_{Fk}^{-1})^{-1}.$$

Collectively, we also denote those $\underline{f}_i^*$'s in (4.6.1.4) as F^*. If the observations are arranged such that the first n_{F1} observations belong to the group 1, the second n_{F2} observations belong to the group 2, and so on, we write

$$Y = [Y_1', Y_2', \ldots, Y_{GF}']',$$

and

$$F = [F_1', F_2', \ldots, F_{GF}']'.$$

The same partition should be used for F^* so that we may write

as well as the estimation procedure for the conditional mode of a subset of the parameters. Unless noted, the use of a globally exchangeable prior distribution for both the factor scores and the factor loadings is assumed.

Conditional Distribution of F Given A, D, H, S, and Y

By collecting L_{Fi} and L_{AFi} from (4.2.14) and (4.3.4), respectively, and using the projection operator trick 2 and the sum of squares trick in the Appendix, we have

(4.6.1.1) $L_{AFi} + L_{Fi}$

$$= \underline{y}_i{}'P\underline{y}_i + (\underline{f}_i - \underline{f}_i^{+})'A'D^{-1}A(\underline{f}_i - \underline{f}_i^{+})$$

$$+ \underline{f}_i{}'C_F^{-1}\underline{f}_i + \ln|C_F|,$$

$$= (\underline{f}_i - \underline{f}^{*})'V^{*-1}(\underline{f}_i - \underline{f}^{*})$$

+ the term not containing $\underline{f}_i$,

where

$$P = I_p - D^{-1}A(A'D^{-1}A)^{-1}A'D^{-1},$$

$$\underline{f}_i^{+} = (A'D^{-1}A)^{-1}A'D^{-1}\underline{y}_i,$$

$$V^{*} = (\ A'D^{-1}A + C_F^{-1}\ )^{-1},$$

and

$$\underline{f}_i^{*} = V^{*}A'D^{-1}\underline{y}_i.$$

Therefore, we have

E_{ADF}

where the factor scores, the error variances and the factor loadings are integrated out in respective order in the E-step.

$A_D E_{AF}$

where the factor scores and the factor loadings are integrated out in respective order in the E-step and the error variances are integrated out analytically after the E-step,

$A_D E_{FA}$

where the factor loadings and the factor scores are integrated out in respective order in the E-step and the error variances are integrated out analytically after the E-step,

$A_{DA} E_F$

where the factor scores are integrated out in the E-step and the factor loadings and the error variances are integrated out analytically in respective order after the E-step.

Point Estimation

In this section the E-step is first described. First, the derivation of E_{DAF}, E_{ADF}, $A_D E_{AF}$, and $A_{DA} E_F$, where the factor scores are first marginalized, is presented. The derivation of E_{DFA} and $E_D E_{FA}$, where the factor loadings are first marginalized, follows. Then, the description of the M-step, where the mode of the hyperparameters are to be estimated, is presented

integrate P2 analytically denoting the result by $f^{*}(H \mid S, Y)$. Also, denote its log by $A_{P2}E_{P1}(H \mid S, Y)$.

The M-step

Maximize $A_{P2}E_{P1}(H \mid S, Y)$ with respect to H.

For the next iteration, P2 should be replaced by the mode or the mean of $f^{*}(P2,H|S,Y)$.

Since this variation is different from the one proposed before the agreement of the results should also support the convergence of those quasi marginalization schemes.

In the application to factor analysis, we partition the parameter P into three subsets, namely,

$P = [F, A, D]$.

Therefore, we have many ways to perform the quasi marginalization by changing the order of integration and the way the parameters are integrated out (by analytical method or taking expectation.) In the following section, only six of them are to be considered. Each of these variations are designated by the particular form which is to be maximized in the M-step. Namely,

E_{DAF}

where the factor scores, the factor loadings and the error variances are integrated out in respective order in the E-step,

E_{DFA}

where the factor loadings, the factor scores and the error variances are integrated out in respective order in the E-step,

the expectation. When the E-step is applied in the next iteration, the first part of E_{P2P1} should be regarded as the complete log likelihood. That is, when calculating the conditional distribution of P1 given P2, H, S and Y, the values of the function of $P2^*$ should be used in place of P2.

The two expectations, E_{h21} and E_{P2P1}, are different since the inclusion of P2 in $P1^*$ is neglected in the latter. Also, note $E_{h12} = E_{h21}$ but $E_{P2P1} <> E_{P1P2}$, where E_{P1P2} denotes the same quasi marginalization scheme with P2 integrated out first. That is, the conditional expectation is affected by the order of integration. Although we cannot prove the convergence of this process to the desired marginal mode, to see that the two different methods, namely, the maximization of E_{P2P1} and the maximization of E_{P1P2}, result in the same solution should support the convergence. If they do converge to the same solution we may as well conclude that the desired marginal mode of H is obtained by this variation of the original EM algorithm.

The quasi marginalization using the EM algorithm was originally suggested by Tom Leonard (personal communication) in the following form.

The E-step

Identify the conditional distribution of P1 given P2, H, S and Y and denote the parameter by $P1^* = h1(P2, H, S, Y)$.

Evaluate the conditional expectation of log complete likelihood with respect to P1 and denote the result by $E_{P1}(P2, H \mid S, Y)$.

Consider $E_{P1}(P2,H|S, Y)$ as the log likelihood of the posterior joint density $f^*(P2,H|S, Y)$ with P1 integrated out and

= Expectation with respect to P1 of Expectation with respect to P2 | P1. However, when the second expectation is taken, it may be the case that it is difficult to do so with $P1^*$ treated as a function of P2, but that it is tractable with $P1^*$ regarded as constant. If this were the case, we propose the following as a variation of the original EM algorithm.

The E-step

Identify the conditional distribution of P1 given P2, H, S and Y and denote the parameter by $P1^* = h1(P2, H, S, Y)$.

Evaluate the conditional expectation of the log complete data likelihood with respect to P1 and denote the result by $E_{P1}(P2, H \mid S, Y)$.

Consider $E_{P1}(P2,H|S, Y)$ as the log likelihood of the posterior joint density $f^*(P2,H|S, Y)$ with P1 integrated out and identify the conditional distribution of P2 given H, S and Y and denote the parameter by $P2^*=h2^*(H,S,Y)$.

Evaluate the conditional expectation of $E_{P1}(P2, H \mid S, Y)$ with respect to P2 given S and Y treating $P1^*$ in the expression as constant and denote the result by $E_{P2P1}(H \mid S, Y)$.

The M-step

Maximize $E_{P2P1}(H \mid S, Y)$ with respect to H.

It is often the case that E_{P2P1} consists of two parts, namely, the part which corresponds to the original complete log likelihood with P1 and P2 replaced by functions of $P1^*$ and $P2^*$, and the additional term due to taking

$$f(P \mid H) = f(P1 \mid H1)\, f(P2 \mid H2).$$

When the posterior marginal mode of

$$g(H \mid S, Y) = \int f(P, H \mid S, Y)\, d\,P$$

$$= \int \int f(P, H \mid S, Y)\, d\,P1\, dP2,$$

are to be estimated we may proceed as follows.

The E-step.

Identify the conditional distribution of P1 given P2, H, S and Y

and denote the parameter by $P1^* = h1(P2, H, S, Y)$.

Identify the conditional distribution of P2 given H, S and Y and

denote the parameter by $P2^* = h2(H, S, Y)$.

Evaluate the conditional expectation of the log complete data likelihood with respect to P1 given P2, H, S and Y and denote the result by $E_{h1}(P2, H \mid S, Y)$.

Evaluate the conditional expectation of $E_{h1}(P2, H \mid S, Y)$

with respect to P2 given H, S and Y noticing that $P1^*$ in the expression is a function of P2 and denote the result by $E_{h21}(H \mid S, Y)$.

The M-step

Maximize $E_{h21}(H \mid S, Y)$ with respect to H

treating $P2^*$ as constant.

This is an authentic application of the EM algorithm in which the expectation is taken successively, i.e.,

Expectation with respect to P1 and P2

= Expectation with respect to P2 of Expectation with respect to P1 | P2

is that they cannot be negative. This can be shown by noticing that RSS's are always non negative and V^* is positive semi definite. Therefore, the solution by the EM algorithm may be different from the one by the direct maximization method such as the one employed in SAS (SAS User's Guide: Statistics, 1982 Edition.) How the algorithm behaves in the neighbourhood of zero is not known.

Also, as many Bayesian statisticians acknowledge, the distributional assumption on [$\underline{a}$, $\underline{b}$] may be thought of as the prior distribution of those effect parameters. (Box (1980), Lindley (1971) Box and Tiao (1973), and Lindley and Smith (1972).) If we accept their view, then, what we have done may be regarded as Parametric Empirical Bayes estimation of the hyperparameters, where the hyperparameters, m, s_A, and s_B are estimated on the basis of data $\underline{y}$. (See Morris (1983).) Also, from the hierarchical Bayes point of view, the solution can be regarded as the marginal posterior mode of the hyperparameters when the hyperparameters, $\underline{m}$ and C, have uniform second stage prior distributions.

Quasi Marginalization by Some Variations of the EM Algorithm

Consider the hierarchical model stated in the previous section, that is,

$$f(P , H \mid Y , S) = f(Y \mid P) f(P \mid H) f(H \mid S),$$

and suppose the partition of the form

$$P = [P1, P2], \quad H = [H1, H2],$$

and

$$+ p \ln(s_A) + q \ln(s_B) + (\underline{a}^* - m\underline{1})'(\underline{a}^* - m\underline{1}) / s_A$$

$$+ (\underline{b}^{*\prime}\underline{b}^*) / s_B + \sum_{i=1}^{p} (v_{ii}) / s_A$$

$$+ \sum_{i=p+1}^{p+q} (v_{ii}) / s_B + \text{const.}$$

The M-step

By differentiating $E_\beta(m, s_A, s_B, s_E | \underline{y})$ with respect to m, s_A, s_B and s_E, and solving the resulting normal equations, we have,

$$m^+ = \sum_{i=1}^{p} (\beta_i^*) / p,$$

$$s_E^+ = (RSS_W + trX'XV^*) / pq,$$

$$s_A^+ = (RSS_A + \sum_{i=1}^{p} (v_{ii})) / p,$$

$$s_B^+ = (RSS_B + \sum_{i=p+1}^{p+q} (v_{ii})) / q,$$

where

$$RSS_W = (\underline{y} - X\underline{\beta}^*)'(\underline{y} - X\underline{\beta}^*),$$

$$RSS_A = \sum_{i=1}^{p} ((a_i^* - m)^2),$$

$$RSS_B = \sum_{i=p+1}^{p+q} (b_i^{*2}).$$

The successive application of these two steps will result in the marginal MLE of s_E, s_A, s_B, and m, with $\underline{\beta}$ integrated out, namely, s_E^+, s_A^+, s_B^+, and m^+, at the convergence point.

One interesting characteristic of those MLE's of the component variances

$$RSS_j^{**} = (\underline{y}_{(j)}-F^*\underline{a}_j^*)'(\underline{y}_{(j)}-F^*\underline{a}_j^*),$$

$$L_{A*} = \sum_{j=1}^{p}[\ L_{Aj*}\ ] + p\ \ln|C_A|,$$

where

$$L_{Aj*} = (\underline{a}_j^*-\underline{a})'C_A^{-1}(\underline{a}_j^*-\underline{a})$$

$$+\ trC_A^{-1}Q_j^*.$$

When the locally exchangeable prior is used for the factor loadings $\underline{Q}_j^*$ and $\underline{a}_j^*$ should be replaced by

(4.6.3.5) $Q_j^* = (\ (1/d_j)F^{*\prime}F^* + C_{Ak}^{-1} + (N/d_j)V^*\ )^{-1},$

and

$$\underline{a}_j^* = Q_j^*(\ (1/d_j)F^{*\prime}\underline{y}_{(j)} + C_{Ak}^{-1}\underline{a}_k),$$

if variable j belongs to the k^{th} locally exchangeable group.

The conditional expectations should be

(4.6.3.6) $E_F(\ A,\ D,\ H\ |\ S,\ Y\ )$ = same as (4.6.3.4)

with L_{A*} replaced by

$$L_{A*} = \sum_{j=1}^{p}[\ L_{Aj*}\ ],$$

where

$$L_{Aj*} = (\underline{a}_j^*-\underline{a}_k)'C_{Ak}^{-1}(\underline{a}_j^*-\underline{a}_k)$$

$$+\ tr\ C_{Ak}^{-1}Q_j^* + \ln|C_{Ak}|,$$

if variable j belongs to the k^{th} group.

Next we consider the analytic marginalization.

$$(4.6.3.7)\quad A_A E_F(\, D,\, H \mid S,\, Y\,) = -2\,\ln(\, \int f^{*}(\, A,\, D,\, H \mid S,\, Y\,)\, d\, A)$$

$$= \sum_{j=1}^{p} [\ \underline{a}_j^{+\prime} F^{*\prime} F^{*} \underline{a}_j^{+} / d_j$$

$$+\ \underline{Y}_{(j)}{}' P \underline{Y}_{(j)} / d_j$$

$$-\ \underline{a}_j^{*\prime} Q_j^{*-1} \underline{a}_j\ ]$$

$$+ \sum_{j=1}^{p} [\ \ln\, |Q_j^{*}|\ ]$$

$$+\ p\ \underline{a}' C_A^{-1} \underline{a} + p\ \ln |C_A| + L_{F*} + L_D + L_H + \text{const..}$$

Now, when $C_A^{-1} \longrightarrow$ zero, this expression reduces to

$$(4.6.3.8)\quad A_A E_F(\, D,\, H, \mid S,\, Y\,)$$

$$= \sum_{j=1}^{p} [\ (\ \underline{Y}_{(j)}{}' F^{*} (F^{*\prime} F^{*})^{-1} F^{*\prime} \underline{Y}_{(j)}$$

$$+\ \underline{Y}_{(j)}{}' P \underline{Y}_{(j)}\ ) / d_j$$

$$-\ \underline{a}_j^{*\prime} Q_j^{*-1} \underline{a}_j^{*}\ ]$$

$$-\ r\ \ln |D| + L_{F*} + L_D + L_H + \text{const},$$

$$= \sum_{j=1}^{p} [\ \underline{Y}_{(j)}{}' \underline{Y}_{(j)} / d_j - \underline{a}_j^{*\prime} Q_j^{*-1} \underline{a}_j^{*}\ ]$$

$$-\ r\ \ln |D| + L_{F*} + L_D + \text{const}.$$

Since, when C_A^{-1} = zero, we have

$$(4.6.3.9)\quad Q_j^{*} = d_j(\ F^{*\prime} F^{*} + N\, V^{*}\ )^{-1},$$

$$\underline{a}_j^{*} = Q_j^{*}(\ (1/d_j) F^{*\prime} \underline{Y}_{(j)}\ )$$

$$= (F^{*\prime}F^{*} + N V^{*})^{-1} F^{*\prime}\underline{Y}_{(j)},$$

$$\underline{a}_j^{*\prime}Q_j^{*-1}\underline{a}_j^{*} = \underline{a}_j^{*}F^{*\prime}\underline{Y}_{(j)}/d_j$$

$$= (\underline{a}_j^{*\prime}(F^{*\prime}F^{*}+N V^{*})\underline{a}_j^{*}) /d_j$$

$$= \underline{a}_j^{*\prime}F^{*\prime}F^{*}\underline{a}_j^{*}+N\underline{a}_j^{*\prime}V^{*}\underline{a}_j/d_j.$$

Therefore,

$$(4.6.3.10)\quad \underline{Y}_{(j)}'\underline{Y}_{(j)}/d_j-\underline{a}_j^{*\prime}Q_j^{*-1}\underline{a}_j^{*}$$

$$= \underline{Y}_{(j)}'\underline{Y}_{(j)}/d_j - 2\underline{a}_j^{*\prime}Q_j^{*-1}\underline{a}_j^{*}$$

$$+ \underline{a}_j^{*\prime}Q_j^{*-1}\underline{a}_j^{*}$$

$$= (\underline{Y}_{(j)}'\underline{Y}_{(j)} - 2\, \underline{a}_j^{*\prime}F^{*\prime}\underline{Y}_{(j)}$$

$$+ \underline{a}_j^{*\prime}F^{*\prime}F^{*}\underline{a}_j^{*} + N\, \underline{a}_j^{*\prime}V^{*}\underline{a}_j^{*})/d_j$$

$$= (RSS_j^{**} + \underline{a}_j^{*\prime}(N V^{*})\underline{a}_j^{*}) / d_j.$$

Finally, we have, when $C_A^{-1} \longrightarrow$ zero,

$$(4.6.3.11)\quad A_AE_F(D, H \mid S, Y)$$

$$= \sum_{j=1}^{p}[(RSS_j^{**} + \underline{a}_j^{*\prime}(N V^{*})\underline{a}_j^{*}) / d_j]$$

$$+ L_{F*} - r \ln|D| + L_D + L_H + \text{const.}$$

Comparing (4.6.3.4) to (4.6.3.11) we notice that in the former expression the additional term due to taking the expectation is added to the summational term and that in the later the multiplicative factor to $\ln|D|$ is reduced by r. In either case this corresponds to the adjustment for the degrees of freedom of

the error variance. See (4.6.9.1) and (4.6.9.2) in the later section. That is, in both cases, the estimate of D will be expanded.

It should also be noted that this method of quasi marginalization does not work when exchangeable prior is used for the factor loadings with nonzero C_A^{-1}.

Quasi Marginalization with respect to the Error Variacnes

Consider, again, that minus twice the log posterior density of D and H is given by $E_{AF}(D,H|S,Y)$ or $A_A E_F(D,H|S,Y)$ and denote it by $f^*(D,H|S,Y)$. (The quasi marginalization of $E_{FA}(D,H|S,Y)$ given in the next section is also considered here.) *Here, again, we may either take the expectation or integrate* D out analytically. We first consider taking the expectation.

Write minus twice the log density as

$$(4.6.4.1)\quad -2 \ln f^*(D,H|S,Y)$$

$$= \sum_{j=1}^{p}[\ (u_j + s)/d_j + (N+n+2) \ln d_j\ ]$$

$$- p(\ n \ln(s/2) - 2 \ln \mathrm{Gamma}(n/2)\ )$$

$$+ L_{F*} + L_{A*} + L_H,$$

where

$$u_j = RSS_j^{**} + tr(F^{*\prime}F^* + N V^*)Q_j^* + N\underline{a}_j^{*\prime} V^* \underline{a}_j^*,$$

if E_{AF} or E_{FA} in the later section is used,

$$u_j = RSS_j^{**} + N\underline{a}_j^{*\prime} V^* \underline{a}_j^*,$$

if $A_A E_F$ is used.

Therefore, the conditional distribution of d_j given H, S, and Y is

(4.6.4.2) $d_j \mid H, S, Y \quad : X^{-2}(N^* + n , u_j + s)$

where $N^* = N$, if E_{AF} or E_{FA} is used, and

$N^* = N - r$, if $A_A E_F$ is used.

Since those distributions are independent for j=1,2,...,p, the joint density is given by the product of the inverted chi square distributions given. We denote the density by $f^*(D|H,S,Y)$.

The conditional expectation is given by

(4.6.4.3) $E_{DAF}(H|S,Y)$ or $E_D A_A E_F(H|S,Y)$

$$= -2 \int \ln f^*(D,H|S,Y) f^*(D|H,S,Y) dD$$

$$= \sum_{j=1}^{p} [(u_j + s)/d_j^* + (N^*+n+2) v_j^*]$$

$$- p(n \ln(s/2) - 2 \ln \mathrm{Gamma}(n/2))$$

$$+ L_{F*} + L_{A*} + L_H + const,$$

where

$$d_j^* = 1/\underline{E}(1/d_j) = (u_j+s)/(N^*+n),$$

and

$$v_j^* = \underline{E}(\ln d_j)$$

$$= \ln((u_j+s)/2) - \mathrm{psy}((N^*+n)/2)$$

where psy(x) is the digamma function, i.e., psy(x) = d lnGamma(x) / dx,

[Johnson and Kotz (1969)].

The term L_{A*} should be subtracted from (4.6.4.3) if ${}_{A}E_{F}$ is used.

Next, consider the analytic integration of D. The integration of $f^{*}(D,H|S,Y)$ with respect to D is straight forward. The result is

(4.6.4.4) $f(H|S,Y) = \prod_{j=1}^{p}[$

$$(s/2)^{n/2}/((u_j+s))^{(N^{*}+n)/2}\text{Gamma}((N^{*}+n)/2)/\text{Gamma}(n/2)\ ].$$

However, instead of straight integration we may approximate the result by using the log normal approximation used in Leonard (1985). That is, the form of $f^{*}(D,H|S,Y)$ suggests that

(4.6.4.5) $u_j \mid d_j \quad : \quad X^2(\ N^{*},\ d_j\ ),$

and

$$d_j \quad : \quad X^{-2}(\ n,\ s\ ),$$

where $X^2(\ n,\ s\ )$ denote the chi square distribution with the degrees of freedom n and the mean ns.

By approximating these distributions by the log normal distributions, Bartlett and Kendall (1946), we have,

(4.6.4.6) $\ln u_j \mid d_j \quad : \quad N(\ \ln d_j + \ln N^{*},\ 2/N^{*}\ ),$

and

$$\ln d_j \quad : N(\ \ln(s/n) + 2/n\ ).$$

Therefore, marginally,

(4.6.4.7) $\ln u_j \quad : \quad N(\ \ln(s/n) + \ln N^{*}\ ,\ 2/N^{*} + 2/n\ ).$

Finally, the result is,

(4.6.4.8) $_{A_D}E_{AF}(H|S,Y)$ or $_{A_{DA}}E_F(H|S,Y)$

$$= \sum_{j=1}^{p} [(\ln(u_j)-(\ln(s/n)+\ln N^*))^2] / (2/N^*+2/n)$$

$$+ L_{F*} + L_H \quad (+ L_{A*}) \quad + \text{const.}$$

When the quasi marginalization of the error variance by taking the expectation is performed prior to the quasi marginalization with respect to the factor loadings we have a similar formula for E_{ADF}, namely,

(4.6.4.9) $E_{ADF}(H|S,Y)$ = same as $E_{DAF}(H|S,Y)$

with $u_j = RSS_j^{**} + N\underline{a}_j^{*}{}'V^*\underline{a}_j^*$,

where RSS_j^{**} is defined in (4.6.3.4),

and for the calculation of all the terms d_j^* should be used.

Given $E_{DAF}(H|S,Y)$, $E_{ADF}(H|S,Y)$, $_{A_D}E_{AF}(H|S,Y)$, or $_{A_{DA}}E_F(H|S,Y)$, we consider those as minus twice the log posterior density of the hyperparameters and the mode of those distributions are to be estimated in the subsequent M-step. However, we first consider the case where the factor loadings are first marginalized.

Conditional Distribution of A Given F, D, H, S, and Y

In this section the derivation of E_{DFA} and $_{A_D}E_{FA}$, where the factor loadings are first marginalized by the EM algorithm will be presented.

By collecting L_{Aj} and L_{FAj} from (4.3.5) and (4.2.16), respectively,

and using the same tricks as before, we have

(4.6.5.1) $L_{FAj} + L_{Aj} =$

$$(\underline{a}_j-\underline{a}_j^*)'Q_j^{*-1}(\underline{a}_j-\underline{a}_j^*) + \text{const},$$

where

$$Q_j^* = (\ (1/d_j)F'F + C_A^{-1})^{-1},$$

$$\underline{a}_j^* = Q_j^*(\ (1/d_j)F'\underline{Y}_{(j)} + C_A^{-1}\underline{a}\ ).$$

This indicates that given F, D, H, S, and Y,

(4.6.5.2) $\underline{a}_j \mid F, D, H, S, Y \quad : \quad N_r(\ \underline{a}_j^*, Q_j^*\ ).$

Since these are independent for $j=1,2,\ldots,p$, the joint density is the product of these densities and we denote it by $f(A|F,D,H,S,Y)$.

When the locally exchangeable prior distribution is used for the factor loadings, similar correction as (4.6.3.5) is necessary. That is, replace all the C_A's and $\underline{a}$'s by C_{Ak} and $\underline{a}_k$ if the variable j belongs to the kth group.

Conditional Expectation of Minus Twice the Log Posterior Density

We have

(4.6.6.1) $E_A(F,D,H|S,Y) = \int L\ f(A|F,D,H,S,Y)\ d\ A$

$$= \sum_{j=1}^{p}[\ (\ RSS_j^* + \text{tr}\ F'FQ_j^*\ )\ /\ d_j + L_{Aj*}\ ]$$

$$+ p\ \ln|C_A| + L_F + L_D + L_H + \text{const.},$$

where

$$RSS_j^* = (\underline{Y}_{(j)}-F\underline{a}_j^*)'(\underline{Y}_{(j)}-F\underline{a}_j^*),$$

$$L_{Aj*} = (\underline{a}_j^* - \underline{a})' C_A^{-1} (\underline{a}_j^* - \underline{a}) + \mathrm{tr} C_A^{-1} Q_j .$$

We write

(4.6.6.2) $L_{A*} = \sum_{j=1}^{p} [\ L_{Aj*}\ ] + p\ \ln|C_A|.$

When the locally exchangeable prior is used similar correction as (4.6.3.6) is necessary.

Quasi Marginalization with respect to the Factor Scores

Now consider $E_A(F,D,H|S,Y)$ as minus twice the log of the posterior density of F,D,H and denote it by $f^*(F,D,H|S,Y)$. Then, by writing observation wise, we have

(4.6.7.1) $E_A(F,D,H|S,Y)$

$$= \sum_{i=1}^{N} [\ (\underline{y}_i - A^* \underline{f}_i)' D^{-1} (\underline{y}_i - A^* \underline{f}_i)$$

$$+ \underline{f}_i' (\ C_F^{-1} + Q./d.\ ) \underline{f}_i\ ]$$

$$+ L_{A*} + L_D + L_H + \text{const.},$$

where

$$Q./d. = \sum_{j=1}^{p} [\ Q_j^* / d_j\ ].$$

Again, using the same tricks, we have

(4.6.7.2) $\underline{f}_i \mid D,H,S,Y \quad : N_r(\ \underline{f}_i^*,\ V^*\ )$

where

$$V^* = (\ A^{*\prime} D^{-1} A^* + Q./d. + C_F^{-1}\ )^{-1},$$

and

$\underline{f}_i^* = V^* A^{*\prime} D^{-1} \underline{y}_i$.

Those conditonal distributions are independent for i=1,2,...,N, the joint density is given by the product of those and we denote it by $f^*(F|D,H,S,Y)$. Also, we write,

(4.6.7.3) $F^* = Y D^{-1} A^* V^*$.

When the locally exchangeable prior is used for the factor scores a similar correction as (4.6.1.4) is necessary.

The conditional expectation of $E_A(F,D,H|S,Y)$ is

(4.6.7.4) $E_{FA}(D,H|S,Y) = \int E_A(F,D,H|S,Y)\ f^*(F|D,H,S,Y)\ dF$

$$= \sum_{i=1}^N [\ (\underline{y}_i - A^* \underline{f}_i^*)' D^{-1} (\underline{y}_i - A^* \underline{f}_i^*)$$

$$+ \underline{f}_i^{*\prime} (\ Q./d. + C_F^{-1}\ ) \underline{f}_i^*\ ]$$

$$+ N\ tr(A^{*\prime} D^{-1} A^* + Q./d. + C_F^{-1}) V^*$$

$$+ L_{F*} + L_{A*} + L_D + L_H + \text{const.}$$

$$= \sum_{j=1}^p [\ (\ RSS_j^{**} + tr(F^{*\prime} F^* + N\ V^*) Q_j^*$$

$$+ N\ \underline{a}_j^{*\prime} V^* \underline{a}_j^*\ ) / d_j\ ]$$

$$+ L_{F*} + L_{A*} + L_D + L_H + \text{const. not including } D,$$

where $RSS_j^{**} = (y_{(j)} - F^* \underline{a}_j^*)' (y_{(j)} - F^* \underline{a}_j^*)$,

and

$$L_{F*} = \sum_{i=1}^N [\ \underline{f}_i^{*\prime} (Q./d. + C_F^{-1}) \underline{f}_i^*]$$

$$+ N\ tr\ C_F^{-1} V^*$$

$+ N \ln|C_F|.$

When the locally exchangeable prior is used for the factor scores a similar correction as (4.6.2.5) is necessary. That is, replace all the occurrences of N times V^* by $\sum_{k=1}^{GF}[n_{Fk}V_k^*]$ and modify L_{F*} to include $\underline{f}_k$.

It should be noted that the final expression in (4.6.7.4), though it is the same as the one in (4.6.3.4), is a different one since the definition of F^* and A^* in them are different. With this difference in mind, we can perform the same quasi marginalization with respect to the error variance as before.

Estimation of the Hyperparameters: the M-step

Here, the minimization of minus twice the conditional expectation of the posterior density of H given S and Y is considered. The expectations are given in (4.6.4.3), (4.6.4.8) and (4.6.4.9). It is now necessary to specify the form of f(H|S). When some information is available we may use the conjugate form suggested in Lindley and Smith (1972) or Lindley (1971). When there are little information we can use uniform prior distribution for H. Although the use of uniform prior distribution in the case where many parameters are to be estimated is subject to criticism, we can argue that it does not hurt the estimation since the number of random variables which have uniform prior distribution is greatly reduced, Lindley (1975). Since the informative specification of the prior distribution of the hyperparameters is usually

difficult, throughout this section uniform prior is used for the hyperparameters.

First, we minimize with respect to the hyperparamters of the factor loadings, H_A. Since the term which includes H_A is the same for all the cases, we have

(4.6.8.1) $\underline{a}^+ = \sum_{j=1}^{p}[\ \underline{a}_j^*\ ]\ /\ p,$

and

$$C_A^+ = (\ A^{*\prime}JA^* + \sum_{j=1}^{p}[\ Q_j^*\ ]\ )\ /\ p,$$

where $J = I_p - (1/p)\underline{1}_p\underline{1}_p'$.

When the globally exchangeable prior is used for the factor scores with $\underline{f}=\underline{0}$ and $C_F=I_r$, those mode are not unique in the sense that an orthogonal rotation of A by an orthonormal matrix, say, T, results in a different mode $T'\underline{a}$ and $T'C_AT$, which also gives the minimum of those expectations. Therefore, the off diagonal elements of C_A should be set equal to zero.

For the hyperparameters of the error varinaces, H_D, when the expectation method is used, we have the following derivatives.

(4.6.8.2)

$$\partial E/\partial n = \sum_{j=1}^{p}[\ v_j^*\ ] - p(\ \ln(s/2) - psy(n/2)\ ),$$

$$\partial E/\partial s = \sum_{j=1}^{p}[\ 1/d_j^*\ ] - pn/s,$$

$$\partial^2 E/\partial n^2 = (1/2)\ p\ psy'(\ n/2\ ),$$

$$\partial^2 E/\partial n\partial s = -\ p/s,$$

$\partial^2 E/\partial s^2 = pn/s^2$,

where psy'(x) is the trigamma function, Johnson and Kotz (1967).

With these derivatives we can solve for n^+ and s^+ by the Newton-Raphson method. Since there are only two parameters involved the process does not require much iterations.

When the log normal approximation is used the solution becomes much more simple. That is,

(4.6.8.3) $n^+ = 1/(\ (1/2)\sum_{j=1}^{p}[(\ln u_j - u.)^2]/p - 1/N^*\ )$

$s^+ = \mathrm{Exp}[\ \sum_{j=1}^{p}[u_j]/p - \ln N^*\ ] \times n^+$

where $u. = \sum_{j=1}^{p}[\ \ln u_j\ ]\ /\ p$.

When the locally exchangeable prior for the factor loadings is used (4.6.8.1) should be replaced by

(4.6.8.4) $\underline{a}_k^+ = \sum_{j \neq k}[\underline{a}_j^*]/n_{Ak}$,

and

$$C_{Ak} = A_k^{*\prime} J_k A_k^* + \sum_{j \neq k}[Q_j^*]/n_{Ak}$$

where the same partition as F^* before is assumed for A^* and J_k is the $n_{Ak} \times n_{Ak}$ centering operator. When a priori zeros are specified for some of the elements of $\underline{a}_k$'s those must be set equal to zero in (4.6.8.4). This does not affect the estimation of C_{Ak}'s since this type of the restrictions of the location parameters can be handled independently of the scale parameteres, Mardia, et.al. (1979).

When the locally exchangeable prior for the factor scores are used we have to estimate the hyperparameters. Differentiation of L_{F*} with correction in (4.6.2.5) gives

(4.6.8.5) $\underline{f}_k^+ = \sum_{i*k}[\underline{f}_i^*]/n_{Fk}$,

and

$$C_{Fk}^+ = F_k^{*\prime} J_k F_k^* / n_{Fk} + V_k^*,$$

where J_k is the $n_{Fk} \times n_{Fk}$ centering operator.

To enforce the orthogonality of the model at least one of the C_{Fk} matrices should be set equal to I_r. Also, to avoid the rotational indeterminacy, one of the C_{Ak}'s should be constrained to a diagonal matrix.

Marginal Mode of the Error Variances Given the Hyperparameters

Conditioned on the hyperparameters, the marginal mode of the error variances can be calculated by a variation of the EM algorithm. That is, depending on the method of quasi marginalization of the factor scores and the factor loadings, the mode of the error variances is given as follows. When E_{AF} or E_{FA} is used, from (4.6.3.4) and (4.6.7.4) we see the marginal mode of the error variance is given by

(4.6.9.1) $d_j^+ = (u_j + s) / (N + n + 2)$, $j=1,2,\ldots,p$,

where

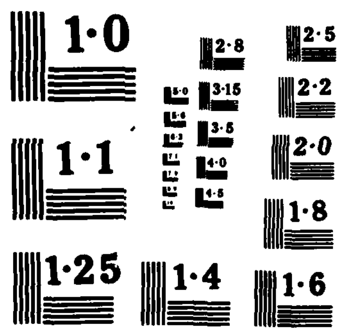
$$u_j = RSS_j^{**} + tr(F^{*\prime}F^* + N V^*)Q_j^* + \underline{a}_j^{*\prime} V^* \underline{a}_j^*$$

AD-A154 184 BAYESIAN FACTOR ANALYSIS(U) IOWA UNIV IOWA CITY 2/3
S I MAYEKAWA 23 MAR 85 TR-85-3-ONR N00014-83-C-0514

UNCLASSIFIED F/G 12/1 NL

END
FILMED
DTIC

cont


1·0
2·8
2·5
3·15
2·2
3·5
1·1
2·0
4·0
4·5
1·8
1·25
1·4
1·6

and RSS_j^{**}, F^*, V^*, A^*, and Q_j^* are defined,

in the corresponding E-step formulae, namely,

either, in (4.6.3.4), (4.6.1.1), and (4.6.3.2) or in (4.6.7.4),

(4.6.7.2), and (4.6.5.1) depending on the choice of E_{AF} or E_{FA},

respectively. When the locally exchangeable proir is used correspoinding corrections should be made.

When $A_A E_F$ is used the estimate becomes, from (4.6.3.11),

$$(4.6.9.2)\quad d_j^+ = (u_j + s) / (N + n - r + 2),\ j=1,2,\ldots,p,$$

where $u_j = RSS_j^{**} + \underline{a}_j^{*\prime}(N V^*)\underline{a}_j^*$.

As noted before, the lack of the term including Q_j^*'s, which comes from taking the expectation with respect to the factor loadings, is compensated by reducing the denominator in (4.6.9.2).

Marginal Mode of the Factor Scores Given the Error Variances

Conditioned on the error variances and the hyperparameters marginal mode of the factor scores can be calculated by the EM algorithm where the factor loadings are treated as the missing data. The result is given in (4.6.5.1), (4.6.6.1), (4.6.7.2), and (4.6.7.3). Here, the first two formulae are considered to be the E-step, and the last two, the M-step. (Due to the symmetry of the conditional density, F^* is not only the mean but also the mode.) It should be noted that this formula is not linear in terms of F^*

since V^* include Q_j^*'s which, in turn include F^*. When the locally exchangeable prior distributions are used corresponding corrections should be made.

Marginal Mode of the Factor Loadings Given the Error Variances

Conditioned on the error variances and the hyperparameters marginal mode of the factor loadings can be calculated by the EM algorithm where the factor scores are treated as the missing data. The result is given in (4.6.1.1-3), (4.6.2.1), and (4.6.3.2). Here, the first four formulae are considered to be the E-step, and the last one, the M-step. (Due to the symmetry of the conditional density, A^* is not only the mean but also the mode.) When the locally exchangeable prior distributions are used corresponding corrections should be made.

Joint Mode of the Factor Scores and the Factor Loadings Given the Error Variances

Conditional on the error variances and the hyperparameters, the joint mode of the factor scores and the factor loadings can be given by a straight forward minimization since there is nothing left for marginalization. Successive application of

(4.6.12.1) $F^+ = Y D^{-1} A V^+$,

where $V^+ = (A'D^{-1}A + C_F)^{-1}$,

and

(4.6.12.2) $\underline{a}_j^+ = Q_j^+((1/d_j)F'\underline{Y}_{(j)} + C_A^{-1}\underline{a}),$

where $Q_j^+ = ((1/d_j)F'F + C_A^{-1})^{-1},$

gives the joint mode. Note that no extra term due to taking the expectation is involved. When locally exchangeable prior distributions are used corresponding corrections should be made.

Summary of the Algorithms

A summary of the six methods propsed in the previous sections for the estimation of the modal value of the hyperparameters is found in Table 1. The necessary formulae are listed for each step, namely, the E-step and the quasi marginalization and the M-step. The two steps are iterated until the process converges.

Technical Notes

Unless specified otherwise the following initial configuration is used.

Initial Factor Loadings

(4.6.14.1) Principal Component Solution based on $(1/N)Y'Y$ matrix.

Initial Error Variances

(4.6.14.2) $\mathrm{diag}[(1/N)Y'Y - A'A]$, where A is the initial factor loadings.

Initial Factor Scores

(4.6.14.3) $Y D^{-1} A (A' D^{-1} A + C_F^{-1})^{-1}.$

When no a priori zeros are specified it may be efficient to orthogonally rotate the initial configuration given above so that

(4.6.14.4) $A'J_pA$ = diagonal

is satisfied and enforce the restriction

(4.6.14.5) C_{Ak} = diagonal

in the M-step. When the hyperparameters have uniform prior distributions this restriction eliminates the rotational indeterminacy. When there are a priori zeros, the following orthogonal rotation to the target matrix B should be performed on the initial configuration given in (4.6.14.1).

(4.6.14.6) $B = [\ b_{je}\ ]$, $j = 1, 2, \ldots, p$, $e = 1, 2, \ldots, r$,

where $b_{je} = 0$ if j belongs to k and $a_{ke} = 0$,

= missing,

where a_{ke}, $k = 1, 2, \ldots, G_A$, $e = 1, 2, \ldots, r$, denotes the locational hyperparameter of the factor loadings.

The iteration process should be terminated when certain convergence criterions is satisfied. For the programs where the calculation is done with double precision numbers the criterion

Successive absolute difference of n is less than or equal to .001,

and

Successive absolute difference of d_j, $j=1,2,\ldots p$, is less than or equal to .00001,

is used. Since the degrees of freedom, n, has a value comparable to the number of observations, the convergence criterion for n should be adjusted according to the number of observations. Also, the convergence criterion for the error

variances should be adjusted according to the observed variance of each variable.

When performing the E_{DFA} or $A_D E_{FA}$ method with a uniform prior distribution of the hyperparameters the conditional expectation of F tends to shrink toward zero. This is due to the unboundedness of the marginal likelihood of F and D in (4.4.1). Therefore, some normalization such as

(4.6.14.7) $F^{*\prime}F^{*} = N\, I_r$,

and corresponding rescaling of V^{*}, C_F, A^{*}, $\underline{a}$, and C_A is necessary. Also, due to the unboundedness of (4.4.3) with uniform prior on s, algorithms such as E_{AFD} or E_{FAD} are impossible. The essential difference between those two cases is that in the latter case, where D is marginalized first, the conditional expectation of $1/d_j$ or d_j is a function of RSS_j only and can easily become zero.

When no a priori zeros are specified we may be able to rotate the solution to a simple structure. This is possible because we have uniform prior distributions for the hyperparameters. It should be noted, however, that the rotation must be performed not only on the parameter matrices but also on the hyperparameters. That is, denoting the rotated matrices by #, the result of the rotation defined in (2.1.14) is

(4.6.14.8) $F^{\#} = F\, T$,

$$A^{\#} = A\,(T')^{-1},$$

$$\underline{a}_k^{\#} = T^{-1}\underline{a}_k, \quad k = 1, 2, \ldots, G_A,$$

$$C_{Ak}^{\#} = T^{-1}C_{Ak}(T')^{-1}, \quad k = 1, 2, \ldots, G_A,$$

$$\underline{f}_k^{\#} = T'\underline{f}_k, \quad k = 1, 2, \ldots, G_F,$$

and $$C_{Fk}^{\#} = T'C_{Fk}T, \quad k = 1, 2, \ldots, G_F.$$

It should be noted that the conditional dispersion matrices such as V^* and Q_j^* may be interpreted as the lower bound of the posterior marginal dispersion matrices of $\underline{f}_i$ or $\underline{a}_j$, respectively, in the sense that the real marginal dispersion matrices are larger than the conditional dispersion matrices. That is, if some of the conditional dispersion matrices are very large it indicates that the factor analytic model with the given dimensionality does not fit the data.

Also, the matrix $(1/N)F^{*\prime}F^*$ is the sample dispersion matrix of the factor scores calculated from the estimate F^*. Therefore, specification errors with respect to the number of dimensions may be checked by the diagonal elements of this matrix. That is, if some of the sample variances are small, it indicates that we might have specified too many factors. In this case the analysis should be performed with a smaller number of factors.

When some of the error variances are zero use

$$\text{(4.6.14.9)} \quad V^* = C_F - C_F A' (D+AC_F A')^{-1} AC_F,$$

$$W = D^{-1}AV^* = (AC_F A' + D)^{-1} AC_F,$$

$$Q_j = d_j(F^{*\prime}F^* + N V^* + d_j C_A^{-1})^{-1},$$

and $$\underline{a}_j = (F^{*\prime}F^* + NV^* + d_j C_A^{-1})^{-1}$$

$$\times (F^{*\prime}\underline{y}_{(j)} + d_j C_A^{-1}\underline{a}).$$

The above expressions given above, which can be derived by using matrix inversion tricks and Lawley's trick, reduce to the expressions given by Rubin and Thayer (1982), namely, (2.2.26) through (2.2.28), when C_A^{-1} is zero.

CHAPTER V

EVALUATION OF THE METHOD

Convergence

In order to check for convergence, the solutions using the following nine methods were compared.

1. M.L.E. by SAS

2. Marginal Estimate of the Eror Variances

   i.e., E_{FA} without marginalization with respect to D

3. Marginal Estimate of the Eror Variances,

   i.e., E_{AF} without marginalization with respect to D.

4. $A_{DA}E_{F}$

5. $A_{D}E_{AF}$

6. $A_{D}E_{FA}$

7. E_{DAF}

8. E_{DFA}

9. E_{ADF}

The three data matrices, namely, sixteen psychological tests from Harman (1976, pp123-124), ten artificial variables from Francis (1983), (see Seber(1984)), and five mathematics tests from Mardia, et., al. (1979) are analyzed with r = 4, 2, and 1, respectively. The first matrix is a correlation

matrix and the rest are dispersion matrices. The elements of the last matrix are divided by 1000 in order to keep the numbers in a reasonable range. For the factor loadings, in order to make the comparison possible, $C_A^{-1} \longrightarrow$ zero is used.

For methods 2 and 3, the convergence criterion is such that the

mean absolute difference of d_j <= .00001,

and for the last six methods, the criterion for convergence require the

absolute difference of n^+ <= .001.

The results are shown in Figures 1-3, where the mean and the variance of the estimated d_j's are also shown. We can conclude from these Figures that six variations of the EM algorithm are almost similar. Therefore, the quasi marginalization may be regarded as a very good approximation to the real marginalization.

Since E_{DAF} is easy to modify for the calculation of MLE and the marginal estimate of the error varinaces in the later analysis only E_{DAF} is used.

Robustness to Initial Configuration

In order to evaluate the robustness or the sensitivity to the initial configuration the correlation matrix in Table 2(e) was analyzed with a variation of the E_{DAF} method with several different initial configurations. The data matrix is calculated by Davis (1944) and used by Martin and McDonald (1975) and found to result in a Heywood case with zero error varinace for the

first variable.

The variables are:

1. Knowledge of word meanings
2. Ability to select the appropriate meaning for a word or phrase in the light of its particular contextual setting
3. Ability to follow the organization of a passage and to identify antecedents and references in it.
4. Ability to select the main thought of a passage
5. Ability to answer questions that are specifically answered in a passage
6. Ability to answer questions that are answered in a passage but not in the words in which the question is asked
7. Ability to draw inferences from a passage about its content
8. Ability to recognize the literary devices used in a passage and to determine its tone and mood
9. Ability to determine a writer's purpose, intent, and point of view, i.e., to draw inference about a writer

A special version of the E_{DAF} program was developed which uses a uniform factor loadings prior, a globally exchangeable factor score prior, and a globally exchangeable error variance prior. The degrees of freedom and the scale parameter of the inverted chi square distributions are calculated with the E_{DAF} program using the same prior distributions as stated above for the factor scores and the factor loadings. Since the main interest here is to see the effect of the globally exchangeable prior distribution of the error variances on the analysis of a Heywood prone data matrix and the comparison

between the Bayesian solutions and the MLE, the marginalizations of the factor loadings and the error variances are not performed. That is, the program calculates the posterior joint mode of the factor loadings and the error variances with the prior distributions specified above. The only difference between this special program and the method proposed by Rubin and Thayer (1982) is that the former uses the inverted chi square prior distributions as the prior distribution of the error variances. It is assumed that the algorithm has converged when the mean partial derivative of (4.4.2) with respect to D is less than or equal to .00001.

Eleven different initial configurations of the factor loadings and the error variances are calculated as follows. (The initial configuration for the factor scores is calculated, given A and D, by (4.6.14.3).)

1. BLR

   Result of E_{DAF} with globally exchangeable factor score prior distribution and uniform factor loading and error variance prior distributions without marginalization of the factor loadings and the error variances. This is the MLE by the EM algorithm proposed by Rubin and Thayer (1982).

2. BLM

   Result of E_{DAF} with globally exchangeable factor score prior distribution uniform factor loading and error variance prior distributions without marginalization of the factor loadings and the error variances. Instead of (4.6.14.2) its mean is used as the initial estimate of all the error variances.

   This is also the MLE by the EM algorithm with a different initial estimate.

3. BHR

Result of E_{DAF} with globally exchangeable factor score and error variance prior distributions and uniform factor loading prior distribution. The hyperparameters n and s are estimated here.

4. BHM

Result of E_{DAF} with globally exchangeable factor score and error variance prior distributions and uniform factor loading prior distribution. Instead of (4.6.14.2) its mean is used as the initial of all the error variances.

Although the hyperparameters are estimated here they are not used in the later analysis.

5. BMR

Result of E_{DAF} with globally exchangeable factor score prior distribution and uniform factor loading and error variance prior distributions without marginalization of the error variances.

This is the marginal MLE of the error variances.

6.

Result of E_{DAF} with globally exchangeable factor score prior distribution and uniform factor loading and error variance prior distributions without marginalization of the error variances.

Instead of (4.6.14.2) its mean is used as the initial of all the the error variances.

This is also the marginal MLE of the error variances.

7. SSMC

MLE by SAS PROC FACTOR with PRIOR SMC option.

$$[\ \partial f/\partial X\ ] = \partial(\mathrm{tr}[\partial f/\partial Y]_c{}'Y)/\partial X.$$

Proof

$$\partial f/\partial x_{pq} = \sum_{i,j}(\ \partial f/\partial y_{ij}\ \partial y_{ij}/\partial x_{pq}\ )$$

$$= \mathrm{tr}\ [\ \partial f/\partial Y\ ]'\ [\ \partial Y/\partial x_{pq}\ ],$$

$$= \partial(\mathrm{tr}[\ \partial f/\partial Y\ ]_c{}'Y)/\partial x_{pq},$$

where $[\ \partial Y/\partial x_{pq}\ ] = [\ \partial y_{ij}/\partial x_{pq}\ ]$, a x b. ||

When evaluating $[\ \partial f/\partial Y\ ]_c$ treat Y as if all of its elements are distinct even if Y is symmetric.

As a special case of b=1 and d=1 we have

$$\partial f/\partial \underline{x} = [\ \partial f/\partial \underline{y}\ ]'\ [\ \partial \underline{y}/\partial \underline{x}\ ],$$

where $[\ \partial \underline{y}/\partial \underline{x}\ ] = [\ \partial y_i/\partial x_j\ ]$, $i=1,2,\ldots,a$, $j=1,2,\ldots,c$, a x c.

Also, when $Y = x\ I$, we have

$$[\ \partial f/\partial x\ ] = \mathrm{tr}[\ \partial \mathrm{tr}[\partial f/\partial Y]_c/\partial x\ ].$$

For some specific forms of f we have

$\partial \mathrm{tr}AX/\partial X = A'$, if X is distinct,

$= A+A'-\mathrm{diag}(A)$, if X is symmetric.

$\partial \mathrm{tr}X'AX/\partial X = (A+A')X$, if X is distinct,

$= (A+A')X+X(A+A')-\mathrm{diag}[(A+A')X]$, if X is symmetric.

$\partial \mathrm{tr}XAX'/\partial X = X(A+A')$, if X is distinct,

$= X(A+A')+(A+A')X-\mathrm{diag}[(A+A')X]$, if X is symmetric.

$\partial \mathrm{tr}X'AXB/\partial X = AXB+A'XB'$, if X is distinct,

$= AXB+BXA+A'XB'+B'XA'-\mathrm{diag}(AXB)-\mathrm{diag}(BXA)$, if X is symmetric.

$\partial X^{-1}A/\partial X = -(X^{-1}A'X^{-1})'$, if X is distinct,

$[\,\partial f/\partial Y\,] = [\,\partial f/\partial y_{ij}\,]$, $i=1,2,\ldots,a$, $j=1,2,\ldots,b$, $a \times b$,

and, the derivative of f with respect to X,

$[\,\partial f/\partial X\,] = [\,\partial f/\partial x_{ij}\,]$, $i=1,2,\ldots,c$, $j=1,2,\ldots,d$, $c \times d$.

Therefore, the derivative with respect to Y' is

$[\,\partial f/\partial Y'\,] = [\,\partial f/\partial Y\,]'$, $b \times a$.

Also, when Y is diagonal,

$[\,\partial f/\partial Y\,] = \mathrm{diag}[\,\partial f/\partial Y\,]$.

Using the following properties of trace we have two usefull rules for differentiation.

$\mathrm{tr}UV = \mathrm{tr}VU = \mathrm{tr}V'U' = \mathrm{tr}U'V'$, and $\mathrm{tr}UV = \sum_{i,j}(\,u_{ij}v_{ji}\,)$.

1. Product Rule

Let U and V be functions of X. Then,

$$\partial \mathrm{tr}UV/\partial X = \partial \mathrm{tr}UV_c/\partial X \quad + \quad \partial \mathrm{tr}U_cV/\partial X,$$

where subscript c denotes that the matrix with c is to be held constant for the purpose of differentiation.

Proof

$$\begin{aligned}
\partial \mathrm{tr}UV/\partial x_{pq} &= \partial \textstyle\sum_{i,j}(\,u_{ij}v_{ji}\,)/\partial x_{pq} \\
&= \textstyle\sum (\partial u_{ij}/\partial x_{pq}\; v_{ji}) + \sum (u_{ij}\; \partial v_{ji}/\partial x_{pq}) \\
&= \partial \textstyle\sum(u_{ij}(v_{ij})_c)/\partial x_{pq} + \partial \sum((u_{ij})_c v_{ji}/\partial x_{pq} \\
&= \partial \mathrm{tr}UV_c/\partial x_{pq} \quad + \partial \mathrm{tr}U_cV/\partial x_{pq}. \;||
\end{aligned}$$

2. Chain Rule

Let Y be a function of X. Then,

$$= \underline{E}\left(\sum_{i,j} (c_{ij} y_{ij}) \right)$$

$$= \sum_{i,j} \left(c_{ij} \underline{E}(y_{ij}) \right)$$

$$= \mathrm{tr} C \underline{E}(Y)$$

$$= \mathrm{tr} C \underline{E}(\underline{x}\underline{x}')$$

$$= \mathrm{tr} C\left(V^* + \underline{E}(\underline{x})\underline{E}(\underline{x})' \right)$$

$$= \mathrm{tr} C\left(V^* + \underline{x}^* \underline{x}^{*\prime} \right)$$

$$= \mathrm{tr} CV^* + \mathrm{tr} C \underline{x}^* \underline{x}'$$

$$= \mathrm{tr} CV^* + \underline{x}^{*\prime} C \underline{x}^* . \quad ||$$

Partial Differentiation of Some Functions of Matrices

Following rules and formulae are collected from Schoenemann (1965), Rao (1973), Press (1982), and Nel (1980).

Definition:

Let $f = f(Y)$,

where f is a matrix to scalar function,

$Y = [y_{ij}]$, $i=1,2,\ldots,a$, $j=1,2,\ldots,b$, $a \times b$,

$X = [x_{ij}]$, $i=1,2,\ldots,c$, $j=1,2,\ldots,d$, $c \times d$,

$y_{ij} = g_{ij}(X)$,

where g_{ij}, $i=1,2,\ldots,a$, $j=1,2,\ldots,b$, is a matrix to scalar function.

Then, the derivative of f with respect to Y is defined as

$$= D^{-1}A(A'D^{-1}A)^{-1}(C_F^{-1}+A'D^{-1}A)-D^{-1}A$$

$$= D^{-1}A(A'D^{-1}A)^{-1}C_F^{-1}.$$

Therefore,

$$\Omega^{-1}A((A'D^{-1}A)^{-1}+C_F)C_F^{-1} = D^{-1}A(A'D^{-1}A)^{-1}C_F^{-1}.$$

$$\Omega^{-1}A((A'D^{-1}A)^{-1}+C_F) = D^{-1}A(A'D^{-1}A)^{-1}.$$

$$\Omega^{-1}A(I+C_FA'D^{-1}A) = D^{-1}A.$$

$$\Omega^{-1}A = D^{-1}A(I+C_FA'D^{-1}A)^{-1}.$$

$$= D^{-1}A(C_F^{-1}+A'D^{-1}A)^{-1}C_F^{-1}. \quad ||$$

Expectation of Quadratic Form

Let $\underline{E}(\underline{x}) = \underline{x}^*$ and $\underline{D}(\underline{x}) = V^*$.

Then $\underline{E}(\ \underline{x}'C\underline{x}\ ) = \underline{x}^{*\prime}C\underline{x}^* + \mathrm{tr}CV^*$.

Proof

$\underline{x}'C\underline{x} = \mathrm{tr}C\underline{x}\underline{x}'$

$$= \textstyle\sum_{i,j}(\ c_{ij}y_{ij}\ ),$$

where $Y = \underline{x}\underline{x}'$.

Therefore,

$\underline{E}(\underline{x}'C\underline{x}) = \underline{E}(\ \mathrm{tr}CY\ )$

Let A, B, and C be, respectively, p x p, p x q, and q x p. Then,

3. $| A + BC | = | A^{-1} | \, | I_p + A^{-1}BC |$

$$= | A^{-1} | \, | I_q + CA^{-1}B |.$$

Lawley's Trick

$$\Omega^{-1}A = D^{-1}A\ (C_F^{-1}+A'D^{-1}A)^{-1}\ C_F^{-1},$$

where $\Omega = A\ C_F\ A' + D$.

Proof

By the Matrix Inversion Formula 1 we have

$$\Omega^{-1} = D^{-1}-D^{-1}A(\ C_F^{-1} + A'D^{-1}A\ )^{-1}A'D^{-1}.$$

Multiply $A(A'D^{-1}A)^{-1}(C_F^{-1}+A'D^{-1}A)$ from the right. Then,

$$\text{left} = \Omega^{-1}A(AD^{-1}A)^{-1}(C_F^{-1}+A'D^{-1}A)$$

$$= \Omega^{-1}A(A'D^{-1}A)^{-1}C_F^{-1} + \Omega^{-1}A$$

$$= \Omega^{-1}A((A'D^{-1}A)^{-1}C_F^{-1}+I)$$

$$= \Omega^{-1}A((A'D^{-1}A)^{-1}+C_F)C_F^{-1}.$$

$$\text{right} = D^{-1}A(A'D^{-1}A)^{-1}(C_F^{-1}+A'D^{-1}A)$$

$$-D^{-1}A(C_F^{-1}+A'D^{-1}A)^{-1}A'D^{-1}A(A'D^{-1}A)^{-1}(C_F^{-1}+A'D^{-1}A)$$

Sum of Squares Completion Trick

Let $\underline{x}$, $\underline{a}$, and $\underline{b}$ be p x 1, and A and B be p x p symmetrix.

Then,

$$q = (\underline{x}-\underline{a})'A(\underline{x}-\underline{a}) + (\underline{x}-\underline{b})'B(\underline{x}-\underline{b})$$

$$= (\underline{x}-\underline{d})'D(\underline{x}-\underline{d}) + c,$$

where,

$$D = A + B,$$

$$\underline{d} = D^{-1}(A\underline{a} + B\underline{b}),$$

and

$$c = \underline{a}'A\underline{a} + \underline{b}'B\underline{b} - \underline{d}'D\underline{d}.$$

Proof

$$q = \underline{x}'(A+B)\underline{x}-2\underline{x}'(A\underline{a}+B\underline{b})+\underline{a}'A\underline{a}+\underline{b}'B\underline{b}$$

$$= (\underline{x}-(A+B)^{-1}(A\underline{a}+B\underline{b}))'(A+B)(\underline{x}-(A+B)^{-1}(A\underline{a}+B\underline{b}))$$

$$-(A\underline{a}+B\underline{b})(A+B)^{-1}(A\underline{a}+B\underline{b})+\underline{a}'A\underline{a}+\underline{b}'B\underline{b}. \quad ||$$

Matrix Inversion and Determinant Trick

Let A and B be square. Then,

1. $(A + UBV)^{-1} = A^{-1} - A^{-1}U(B^{-1} + VA^{-1}U)^{-1}VA^{-1}.$
2. $(A + B)^{-1} = A^{-1} (A^{-1} + B^{-1})^{-1} B^{-1}.$

But,

$$Y^{+} - XB = XB^{+} - XB = X(\ B^{+}-B\ ),$$

and

$$U^{+\prime}(Y^{+}-XB) = Y'PX(B^{+}-B) = \text{zero},$$

since PX = zero. ||

Projection Operator Trick 2

Let Z be $n \times p$, V, $n \times r$, B, $r \times p$, and A, $p \times p$.

Then,

$$Q = (\ Z - VB\ )'\ A^{-1}\ (\ Z - VB\ )$$

$$= Z'A^{-1/2}PA^{-1/2}Z + (\ B - B^{+}\ )'\ V'A^{-1}V\ (\ B - B^{+}\ ),$$

where,

$$P = I_n - A^{-1/2}V\ (V'A^{-1}V)^{-}\ V'A^{-1/2},$$

and

$$B^{+} = (\ V'A^{-1}V\ )^{-}V'A^{-1}Z.$$

Proof

Apply the trick 1 to the transformed variables

$Y = A^{-1/2}Z$ and $X = A^{-1/2}V$. ||

APPENDIX

Projection Operator Trick 1

Let Y be n x p, X, n x r, and B, r x p.

Then,

$$Q = (Y - XB)'(Y - XB)$$

$$= Y'PY + (B - B^{+})' X'X (B - B^{+}),$$

where

$$P = I_{n} - X (X'X)^{-} X,$$

$$B^{+} = (X'X)^{-}X'Y,$$

and

A^{-} is a generalized inverse of A.

Proof

Let $Y^{+} = X B^{+} = X(X'X)^{-}X'Y$,

and $U^{+} = Y - Y^{+} = PY$.

Then,

$$Q = (Y-Y^{+}+Y^{+}-XB)'(Y-Y^{+}+Y^{+}-XB)$$

$$= (U^{+} + Y^{+}-XB)'(U^{+} - Y^{+}-XB)$$

$$= U^{+\prime}U^{+} + U^{+\prime}(Y^{+}-XB) + (Y^{+}-XB)'U^{+} + (Y^{+}-XB)'(Y^{+}-XB).$$

the later stage. As we expect, the modal values are slightly larger than the expected values. Next, we observe more shrinkage toward the mean in the marginal estimate. This can be checked by the sample variances of each set of the parameters. This is also expected since the further we marginalize the smaller the conditional dispersion matrix becomes. For the factor scores compare V's in (4.6.12.1) and (4.6.7.3) and for the factor loadings, Q's in (4.6.12.2) and (4.6.3.2).

As far as this data set is concerned, it can be said that the choice of a particular set of the estimates is purely subjective.

Conclusions

It is found that six methods to marginalize parameters behave very similarly. Therefore, the use of the simplest and the most natural method, namely, the E_{DAF} method, is recommended. It is also found that the method is robust to the choice of initial estimates. Finally, use of locally exchangeable prior distribution for the factor loadings is highly recommended when there are some grouping information of the variables prior to the analysis.

3. Flags
4. General Information
5. Paragraph Comprehension
6. Sentence Completion
7. Word Classification
8. Word Meaning
9. Addition
10. Code
11. Counting Dots
12. Straight-Curved Capitals
13. Word Recognition
14. Number Recognition
15. Object-Number
16. Number-Figure

Since it is known that the variables form four clusters, [Shiba (1979)], the following four locally exchangeable groups with a priori zeros are used.

Group 1. Variables 1, 2, and 3, with zeros in dimensions 2, 3, and 4.

Group 2. Variables 4, 5, 6, 7, and 8, with zeros in dimensions 1, 3, and 4.

Group 3. Variables 9, 10, 11, and 12, with zeros in dimensions 1, 2, and 4.

Group 4. Variables 13, 14, 15, and 16, with zeros in dimensions 1, 2, and 3.

With this prior specification the data set was analyzed with the E_{DAF} method and the results of each stage are shown in Tables 6(a) through 6(h).

We first notice that the difference between the first stage and the second stage is not large. Therefore, it may be reasonable to skip the second stage and use the conditional expectation of (reciprocal of) the error variance in

The estimated hyperparameters are shown in Table 5(g).

Choice of Estimates

As noted before, the choice of the estimates depends on the investigator's interest. If we are interested in both the factor scores and the factor loadings, the joint mode sould be used. However, if we are interested in one of them the marginal mode should be used. Also, in each stage of the estimation procedure, that is,

Stage 1. Marginal Mode of the Hyperparameters

Stage 2. Marginal Mode of the Error Variances Conditional on the Hyperparameters

Stage 3. Joint Mode of the Factor Scores and the Factor Loadings Conditional on the Error Variances and the Hyperparameters

Stage 4. Marginal Mode of the Factor Loadings Conditional on the Error Variances and the Hyperparameters

Stage 5. Marginal Mode of the Factor Scores Conditional on the Error Variances and the Hyperparameters

all the parameter values are available, if not as the mode, as the conditional expectation. Therefore, it may be reasonable not to perform all the analysis if those parameter values are similar.

In order to check this aspect of the procedure the correlation matrix in Table 2(a) was analyzed. The variables used are:

1. Visual Perception
2. Paper Form Board

9. Mechanical Comprehension

10. Electronics Information

The data were analyzed by the A_DE_{AF} method with three different prior distributions of factor loadings, namely, the locally exchangeable prior with a priori zeros in the hyperparameters, the locally exchangeable prior with the same groupings but without *a priori* zeros, and the globally exchangeable prior. The grouping of the variables and the locations of the a priori zeros are

Group 1: Variables 1,3 and 4, zeros in Dimensions 2,3, and 4.

Group 2: Variables 7,9, and 10, zeros in Dimensions 1,3, and 4.

Group 3: Variables 2 and 8, zeros in Dimensions 1,2, and 4.

Group 4: Variables 5 and 6, zeros in Dimensions 1,2, and 3.

These grouping and a priori zeros are suggested by the analysis by Ree, et.al. (1981). The globally exchangeable prior distribution of the error variance is used and the hyperparameters, n and s, are estimated first. Then, conditioned on those values and the final value of D, the marginal mode of the factor loadings are calculated. The marginal mode of D is not used since, at least with this data, it is very similar to the final value of D.

The results are shown in Tables 5(a) through 5(d) with an additional solution from SAS PROC FACTOR MLE with PRIOR SMC option, Table 5(e), and the oblique solution in Ree, et.al. (1981), Table 5(f). The results in Tables 5(b) through 5(e) are rotated by the VARIMAX method. By comparing Table 5(a) and 5(b) it is obvious that just by specifying a priori grouping and zeros in the hyperparameters we can attain simple structure. Ironically, in this case, the VARIMAX solution seems to be worse than the unrotated solution. The rest of the Table 5 also shows the same pattern of simple structure.

It can be concluded that the method is very robu ` to the different initial configurations. Also this finding strongly demonstrates the superiority of the Bayesian factor analysis proposed in the Chapter IV over the usual MLE. Five different MLE's found by SAS PROC FACTOR are reduced to two different Bayesian solutions just by assuming an informative prior distribution of the error variances. Intuitively, it can be said that the prior distribution has the effect of deemphasizing local minima. A similar result may be expected from the use of the prior distributions of the factor loadings.

Effect of the Locally Exchangeable Prior Distributions of the Factor Loadings

In order to demonstrate the effect of the locally exchangeable prior distributions of the factor loadings the correlation matrix of ASVAB Form 8a in Table 2(e) is analyzed. The data were collected by Ree, Mullins, Mathews, and Massey (1981) with 2620 subjects and consists of scores on the following ten tests.

1. General Science
2. Arithmetic Reasoning
3. Word Knowledge
4. Paragraph Comprehension
5. Numerical Operations (speeded)
6. Coding Speed (speeded)
7. Auto-Shop Information
8. Mathematics Knowledge

8. SBLR

MLE by SAS PROC FACTOR with PRIOR uniquenesses given by the result of BLR.

9. SX#2

MLE by SAS PROC FACTOR with the following PRIOR uniquenesses

.5 .001 .5 .5 .5 .5 .5 .5 .5

10. SX#7

MLE by SAS PROC FACTOR with the following PRIOR uniquenesses

.5 .5 .5 .5 .5 .5 .001 .5 .5

11. SX#9

MLE by SAS PROC FACTOR with the following PRIOR uniquenesses

.5 .5 .5 .5 .5 .5 .5 .5 .001

The SAS results with the PRIOR uniquenesses given by the results of BLM, BHR, BHM, BMR, and BMM are the same as SSMC. The initial configurations are shown in Table 3.

Each initial configuration above was submitted to the special program to check the robustness. The values n = 15.0139 and s = 5.48453 were used for all the calculations. As shown in Fig. 4 and Tables 4(a) through 4(d), the program resulted in only two different solutions. The first group, where variable 4 has low error variance, consists of BKBLR, BKBHR, BKBHM, and BKSBR, where BK is attached to indicate the results of the special program. The second group, where variable 4 has high error variance, consists of BKBLM, BKBMR, BKBMM, BKSSMC, BKSBLM, BKSBHR, BKSBHM,BKSBMR, BKSBMM, BKSX#2, BKSX#7 and BKSX#9. The solutions within each group are identical to each other to six digits.

$$= -2\ X^{-1}AX^{-1} - \mathrm{diag}(X^{-1}AX^{-1}),\ \text{if X and A are symmetric.}$$

$$\partial Y^{-1}A/\partial X = -\partial\mathrm{tr}[(Y^{-1}AY^{-1})_c Y]/\partial X.$$

$$\partial \ln|X|/\partial X = (X^{-1})',\ \text{if X is distinct,}$$

$$= 2X^{-1} - \mathrm{diag}(X^{-1}),\ \text{if X is symmetric.}$$

$$\partial \ln|Y|/\partial X = \partial\mathrm{tr}[\ Y_c^{-1}Y\ ]/\partial X.$$

The Guttman/Kestelman Formula

Let the following factor analytic model hold.

$$\underline{y} = A\underline{f} + D\underline{z},$$

where $\underline{y}$ is p x 1, $\underline{f}$ is r x 1, $\underline{z}$ is p x 1,

A is p x r, and D is p x p diagonal,

$\underline{E}(\underline{f}) = \underline{E}(\underline{z}) = \underline{0}$,

$\underline{D}(\underline{f}) = I_r$, $\underline{D}(\underline{z}) = I_p$,

$\mathrm{Cov}(\underline{f},\underline{z})$ = zero.

These imply

$\underline{E}(\underline{y}) = \underline{0}$, and $\underline{D}(\underline{y}) = AA' + D^2$ =, say, Ω.

Note that the error variances are denoted by D^2.

Kestelman(1952) and Guttman(1955) showed that, given A and D, the model is not unique in the sence that the variables

$$\underline{f}^* = A'\Omega^{-1}\underline{y} + P\underline{s},$$

and

$$\underline{z}^* = D\Omega^{-1}\underline{y} - D^{-1}AP\underline{s},$$

where

$$PP' = I_r - A'\Omega^{-1}A,$$

and $\underline{s}$ is any variable which satisfies

$$\underline{E}(\underline{s}) = \underline{0},\ \underline{D}(\underline{s}) = I_p,\ \text{and}\ \mathrm{Cov}(\ \underline{y},\underline{s}\ ) = \text{zero},$$

also satisfy the model requirement stated above.

Proof

$$\underline{y}^* = A\underline{f}^* + D\underline{z}^*$$

$$= A(A'\Omega^{-1}\underline{y}+P\underline{s}) + D(D\Omega^{-1}\underline{y}-D^{-1}AP\underline{s})$$

$$= AA'\Omega^{-1}\underline{y} + AP\underline{s} + D^2\Omega^{-1}\underline{y} - AP\underline{s}$$

$$= (AA'+D^2)\Omega^{-1}\underline{y}$$

$$= \underline{y}.$$

$$\underline{E}(\underline{f}^*) = A'\Omega^{-1}\underline{E}(\underline{y}) + P\underline{E}(\underline{s})$$

$$= \underline{0}.$$

$$\underline{E}(\underline{z}^*) = D\Omega^{-1}\underline{E}(\underline{y}) + D^{-1}AP\underline{E}(\underline{s})$$

$$= \underline{0}.$$

$$\underline{D}(\underline{f}^*) = A'\Omega^{-1}\ \Omega\ \Omega^{-1}A + P\ I\ P'$$

$$= A'\Omega^{-1}A + I - A'\Omega^{-1}A$$

$= I.$

$$\underline{D}(\underline{z}^*) = D\Omega^{-1}\ \Omega\ \ \Omega^{-1}D + D^{-1}AP\ I\ \ P'A'D^{-1}$$

$$= D\Omega^{-1}D + D^{-1}A(I-A'\Omega^{-1}A)A'D^{-1}$$

$$= D^{-1}(D^2\Omega^{-1}D^2+AA'-AA'\Omega^{-1}AA')D^{-1}$$

$$= D^{-1}(D^2\Omega^{-1}D^2+AA'-(\Omega-D^2)\Omega^{-1}AA')D^{-1}$$

$$= D^{-1}(D^2\Omega^{-1}D^2+AA'-AA'+D^2\Omega^{-1}AA'))D^{-1}$$

$$= D^{-1}(D^2\Omega^{-1}D^2+D^2\Omega^{-1}(\Omega-D^2))D^{-1}$$

$$= D^{-1}(D^2\Omega^{-1}D^2+D^2-D^2\Omega^{-1}D^2)D^{-1}$$

$$= D^{-1}(\Omega-AA')D^{-1}$$

$$= I.$$

$$\mathrm{Cov}(\underline{f}^*,\underline{z}^*) = \mathrm{Cov}(A'\Omega^{-1}\underline{y}, D\Omega^{-1}\underline{y}-D^{-1}AP\underline{s})$$

$$= A'\Omega^{-1}\ \Omega\ \Omega^{-1}D - PP'A'D^{-1}$$

$$= A'\Omega^{-1}D - (I-A'\Omega^{-1}A)A'D^{-1}$$

$$= A'(\Omega^{-1}D-D^{-1}+\Omega^{-1}AA'D^{-1})$$

$$= A'(\Omega^{-1}D-D^{-1}+\Omega^{-1}AA'D^{-1})\ DD^{-1}$$

$$= A'(\Omega^{-1}D^2-I+\Omega^{-1}AA')\ D^{-1}$$

$$= A'(\Omega^{-1}(D^2+AA')-I)\ D^{-1}$$

$$= A'(\Omega^{-1}\Omega-I)\ D^{-1}$$

$= \text{zero.}$ ||

BIBLIOGRAPHY

Anderson, T. W. An Introduction to Multivariate Statistical Analysis Second Edition, John Wiley and Sons, 1984, New York.

Anderson, T. W., Rubin, H. (1956). Statistical inference in factor analysis, In J. Neyman Ed. Proc. of Third Berkeley Symposium on Mathematical Statistics and Probability 1956, 5, 111-150.

Bartholomew, D. J. Posterior analysis of the factor model. British J. of Math. and Stat. Psychol., 1981, 34, 93-99.

Bartlett, M. S., Effect of standardization on a X^2 approximation in factor analysis. Biometrika, 1951, 38, 337-344.

Bartlett, M. S., and Kendall, D. G. The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation. JRSS, 1946, 8, 128-138.

Bechtel, G. Multidimensional Preference Scaling The Hague: Mouton, 1976.

Bock, R. D., and Aitkin, M., Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. Psychometrika, 1981, 46, 443-459.

Box, G. E. P. Sampling and Bayes' Inference in Scientific Modeling and Robustness. J. R. Stat. Soc. Series A, 1980, 143, 383-430.

Box, G. E. P., and Tiao, G. E. Bayesian Inference in Statistical Analysis. Reading, Mass., Addison-Wesley, 1973.

Carroll, D. J. Individual differences and Multidimensional Scaling. In Shepard, R. N., Romney, A. K., and Nerlove S. B. ed. Multidimensional Scaling Vol I, Theory. New York: Seminar Press. 1972.

Davis, F. B. Fundamental Factors of Comprehension in Reading. Psychometrika, 1944, 9, 185-197.

Dempster, A. P., Laird, N. M. and Rubin, D. B. Maximum Likelihood from Incomplete Data via the EM Algorithm. J. Royal. Stat Soc, Series B, 1977, 39, 1-38.

Eckart, C. and Young, G. The Approximation of One Matrix by Another of Lower Rank. Psychometrika, 1936, 1, 211-218.

Feldt, L. S. Estimation of the Reliability of a Test Divided into Two Parts of Unequal Length. Psychometrika, 1975, 40, 557-561.

Fienberg, S. F. Discussion of a paper by Professor Lindley and Dr Smith. J. R. Stat. Soc., 1972, 34, 30-32.

Francis, I., Factor Analysis: Its Purpose, Practice, and Pakage Program. Invited Paper, American Statistical Association, New York, Dec., 1983.

Guttman, L., The determinacy of factor score matrix with implication for five other basic problems of common factor theory. BrJ Stat. Psych. 1955, 8, 65-81.

Harman. H. H., Modern Factor Analysis, 3rd edition. The University of Chicago Press, Chicago, Il., 1976.

Herz C. S. Bessel Function of Matrix Argument. Ann. Math. 1955. 61 No.3, 474-523.

Joereskog, K. G., Statistical Estimation in Factor Analysis. A new technique and its foundation. Stockholm: Almqvist β Wiksell, 1963.

Joereskog, K. G. Some contributions to maximum likelihood factor analysis. Psychometrika, 1967, 32, 443-482.

Joereskog, K. G. (1969) A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 34, 183-202.

Joereskog, K. G., Factor analysis by least squares and maximum likelihood methods. In Statistical methods for Digital Computers, ed. K. Enslein, A. Ralston and H. S. Wilf. New York: Wiley, 125-153, 1979.

Joereskog, K. G. and Soerbom, D. (1979). Advances in Factor Analysis and Structural Equation Models. Cambridge, Mass.: Abt Books.

Johnson, N. L., and Kotz, S. Distribution in Statistics: Discrete Distributions, Houghton Mifflin Co., Boston, 1969.

Kaufman, G. M. and Press, S. J. Bayesian factor analysis. Report 7322. University of Chicago. 1973.

Kristof, W. M. Estimation of Reliability and True Score Variance from a split of a test into tree arbitrary parts. Psychometrika, 1974, 39, 491-499.

Lawley, D. N., Further investigation in factor estimation. Proc. Roy. Soc. Edin., 1942, 61, 176-185.

Lawley, D. N. and Maxwell, A. E., Factor Analysis as a Statistical Method, London, Butterworth. 1963.

Lawley, D. N. and Maxwell, A. E., Factor Analysis as a Statistical Method. 2nd ed. London, Butterworth, 1971.

Lee, S. Y. A., Bayesian approach to confirmatory factor analysis. Psychometrika, 1981, 46.

Leonard, T., An Improved Method in m-group Regression. (in preparation.)

Lindley, D. V. Estimation of Many Parameters. In V.P.Godambe and D.A. Sprott, Eds. Foundations of Statistical Inference Statistical Inference, pp435-455. Toronto: Holt, Rinehart and Winston. 1971a.

Lindley, D. V. Bayesian Statistics, A Review. Philadelphia, SIAM., 1971b.

Lindley, D. V. and Smith A. F. M. Bayes Estimates for the Linear Model. J. R. Statist. Series B, 1972, 31,1-41.

Lord, F. M. and Novick, M. R., Statistical Theories of Mental Test Scores. Reading, Mass: Addison-Wesley. 1968.

Mardia, K. V., Kent, J. T., and Bibby, J.M. Multivariate Analysis. London: Academic Press. 1979.

Martin, J. K. and McDonald, R. P. (1975). Bayesian estimation in unrestricted factor analysis: a treatment for Heywood cases. Psychometrika. 1975, 40, 5050-517.

McDonald, R. P., and Burr, E. J. A Comparison of Four Methods of Constructing Factor Scores. Psychometrika, 1967, 32, 381-401.

McDonald, R. P. The structural analysis of multivariate data: A sketch of a general theory. Multivariate Behavioral Research. 1979a, 14, 21-38.

McDonald, R. P., The simultaneous estimation of factor loadings and scores. BrJ Math. Stat. Psychol.. 1979b, 32, 212-228.

Morris, C. N. Parametric Empirical Bayes Inference: Theory and Applications. J.A.S.A., 1983, 78, 47-55.

Mulaik, S. A. The Foundations of Factor Analysis. New York: McGraw-Hill Book Co. 1972.

Nel, D. G., On Matrix Differentiation, South African Stat. Jour., 1980, vol. 14, p137-193.

Novick, M. R., Jackson, P. H., Thayer, D. T., and Cole, N. S. Estimating Multiple Regressions in m-group: a cross validation study. Brt.J.Math.Stat.Psychol., 1972.

O'Hagan, A. On Posterior Joint and Marginal Modes, Biometrika, 1976,63, p 329-233.

Orchard, T. and Woodbury, M. A., A Missing Information Principle: Theory and Applications. Proc. 6th Berkeley Symposium on Math Stat and Prob., 1,1972, 697-715.

Press, S. J. Applied Multivariate Analysis. New York: Holt, Rinehart and Winston, 1972.

Press, S. J. Applied Multivariate Analysis: Using Bayesian and Frequentist Method of Inference. Second Edition. Malabar, Florida: Robsert E. Krieger Publishing Co. 1982.

Rao, C. R. Linear Statistical Inference and Its Application 2nd ed. New York: Wiley. 1973.

Ree, M., J., Mullins, C., J., Mathews, J., J., and Massey, R., H. Armed Services Vocational Aptirude Battery: Item and Factor Analysis of Forms 8, 9, and 10. AFRL Technical Report 81-55, Brooks AFB, Tx.

Rubin, D. B. and Thayer, D. T. EM Algorithm for ML Factor Analysis. Psychometrika, 1982, 47, 69-76.

SAS Institute, SAS User's Guide: Statistics, Cary, North Carolina, 1982.

Schonemann, P. H., On the Formal Differentiation of Traces and Determinants. Research Memorandom. #27, University of North Carolina, 1965.

Seber, G. A. F., Multivariate Observations, 1984, John Wiley and Sons, N.Y.

Shiba, S. New estimats of Factor Scores. Jap. Psychol. Res. 1969. 11, 129-133.

Shiba, S. Inshibunsekiho. (Factor Analysis), 2nd Ed., University of Tokyo Press, Tokyo, 1979. (in Japanese.)

Spearman, C. General intelligence, objectively determined and measured. American Journal of Psychology, 1904, 15, 201-293.

Thomson, G. H. Factor Analysis of Hunam Ability London Univ. Press, London, 1951.

Whittle, P. On principal Components and Least Squares Methods of Factor Analysis. Skand. Aktuar. 1952. 35, 223-239.

Wilson, E. B. Comments on professor Spearman's note. Journal of Educational Psychology, 1929, 20, 217-223.

Wong, G. A Bayesian Approach to Factor Analysis. Program Statistics Research Technical Report. ETS., 1980. No.80-9.

Wu, C., F., On the Convergence Properties of the EM Algorithm, Annals of Stat, 1983, 11, pp95-103.

Table 1 Summary of the Algorithm

	Prior Distributions of F and A	
	Globally Exchangeable	Locally Exchangeable
E_{DAF}		
The E-step		
V^*	4.6.1.1	4.6.1.4
$Y'F^*$	4.6.2.2	4.6.2.6
$F^{*\prime}F^*$	4.6.2.3	4.6.2.7
A^*	4.6.3.2	4.6.3.5
Q^*	4.6.3.2	4.6.3.5
D^*	4.6.4.3	
The M-step		
$\underline{a}$	4.6.8.1	4.6.8.4
C_A	4.6.8.1	4.6.8.4
$\underline{f}$	fixed	4.6.8.5
C_F	fixed	4.6.8.5
n	4.6.8.2 with $\underline{u}$ and N^* in 4.6.4.1 and 4.6.4.2	
E_{DFA}		
The E-step		
A^*	4.6.5.1	4.6.3.5*
Q^*	4.6.5.1	4.6.3.5*
V^*	4.6.7.2	4.6.2.5*
$Y'F^*$	4.6.2.2	4.6.2.6
$F^{*\prime}F^*$	4.6.2.3	4.6.2.7
D^*	4.6.4.3	
The M-step		
$\underline{a}$	4.6.8.1	4.6.8.4
C_A	4.6.8.1	4.6.8.4
$\underline{f}$	fixed	4.6.8.5
C_F	fixed	4.6.8.5
n	4.6.8.2 with $\underline{u}$ and N^* in 4.6.4.1 and 4.6.4.2	

'*' after the fomula number indicates analogous correction.

Table 1 (continued)

	Prior Distributions of F and A	
	Globally Exchangeable	Locally Exchangeable
E_{ADF}		
The E-step		
V^*	4.6.1.1	4.6.1.4
$Y'F^*$	4.6.2.2	4.6.2.6
$F^{*\prime}F^*$	4.6.2.3	4.6.2.7
D^*	4.6.4.9	
A^*	4.6.3.2	4.6.3.5
Q^*	4.6.3.2	4.6.3.5
The M-step		
$\underline{a}$	4.6.8.1	4.6.8.4
C_A	4.6.8.1	4.6.8.4
$\underline{f}$	fixed	4.6.8.5
C_F	fixed	4.6.8.5
n	4.6.8.2 with $\underline{u}$ and N^* in 4.6.4.9 and 4.6.4.2	
$A_D E_{AF}$		
The E-step		
V^*	4.6.1.1	4.6.1.4
$Y'F^*$	4.6.2.2	4.6.2.6
$F^{*\prime}F^*$	4.6.2.3	4.6.2.7
A^*	4.6.3.2	4.6.3.5
Q^*	4.6.3.2	4.6.3.5
The M-step		
$\underline{a}$	4.6.8.1	
C_A	4.6.8.1	
n	4.6.8.3 with $\underline{u}$ and N^* in 4.6.4.1 and 4.6.4.2	

For the next iteration use

$$d_j = (u_j + s) / (N + n + 2).$$

Table 1 (continued)

	Prior Distributions of F and A Globally Exchangeable	Locally Exchangeable
$A_D E_{FA}$		
The E-step		
A^*	4.6.5.1	4.6.3.5*
Q^*	4.6.5.1	4.6.3.5*
V^*	4.6.7.2	4.6.2.5*
$Y'F^*$	4.6.2.2	4.6.2.6
$F^{*\prime}F^*$	4.6.2.3	4.6.2.7
The M-step		
$\underline{a}$	4.6.8.1	4.6.8.4
C_A	4.6.8.1	4.6.8.4
$\underline{f}$	fixed	4.6.8.5
C_F	fixed	4.6.8.5
n	4.6.8.3 with $\underline{u}$ and N^* in 4.6.4.1 and 4.6.4.2	

For the next iteration use

$$d_j = (u_j + s) / (N + n + 2).$$

$A_{DA}E_F$		
The E-step		
V^*	4.6.1.1	4.6.3.4
$Y'F^*$	4.6.2.2	4.6.2.6
$F^{*\prime}F^*$	4.6.2.3	4.6.2.7
A^*	4.6.3.9	
Q^*	4.6.3.9	
The M-step		
$\underline{f}$	fixed	4.6.8.5
C_F	fixed	4.6.8.5
n	4.6.8.3 with $\underline{u}$ and N^* in 4.6.4.1 and 4.6.4.2	

For the next iteration use

$$d_j = (u_j + s) / (N + n - r + 2).$$

Table 2 Data Matrices

(a) Correlation Matrix from Harman (1976)

N = 145

1.000										
.403	1.000									
.468	.305	1.000								
.321	.247	.227	1.000							
.335	.268	.327	.622	1.000						
.304	.223	.335	.656	.722	1.000					
.332	.382	.391	.578	.527	.619	1.000				
.326	.184	.325	.723	.714	.685	.532	1.000			
.116	.075	.099	.311	.203	.246	.285	.170	1.000		
.308	.091	.110	.344	.353	.232	.300	.280	.484	1.000	
.314	.140	.160	.215	.095	.181	.271	.113	.585	.428	1.000
.489	.321	.327	.344	.309	.345	.395	.280	.408	.535	.512
1.000										
.125	.177	.066	.280	.292	.236	.252	.260	.172	.350	.131
.195	1.000									
.238	.065	.127	.229	.251	.172	.175	.248	.154	.240	.173
.139	.370	1.000								
.176	.177	.187	.208	.273	.228	.255	.274	.289	.362	.278
.194	.341	.345	1.000							
.368	.211	.251	.263	.167	.159	.250	.208	.317	.350	.349
.323	.201	.334	.448	1.000						

(b) Dispersion Matrix from Francis (1983).

N = 50

10.934									
8.104	10.709								
10.468	8.623	14.528							
8.541	10.155	9.629	14.846						
11.998	10.494	12.625	11.063	19.832					
−0.047	0.025	0.901	1.892	−0.092	15.804				
2.385	2.765	3.480	3.880	2.915	12.921	17.580			
−0.626	−0.166	0.867	1.466	0.684	15.001	15.426	24.436		
−0.168	1.990	1.433	3.292	0.323	13.498	15.365	16.602	21.326	
−1.749	−1.139	−0.453	1.873	−2.958	12.491	13.367	15.814	15.385	22.136

(c) Dispersion Matrix from Mardia, et.al. (1979).

N = 100

.3023				
.1258	.1709			
.1004	.0842	.1116		
.1051	.0936	.1108	.2179	
.1161	.0979	.1205	.1538	.2944

All the entries are divided by 1000.

Table 2 (continued)

(d) Correlation Matrix of ASVAB Form 8a.

N = 2620

1.00									
.71	1.00								
.83	.70	1.00							
.74	.70	.82	1.00						
.48	.59	.52	.55	1.00					
.43	.52	.48	.49	.64	1.00				
.70	.60	.68	.63	.40	.42	1.00			
.65	.79	.62	.60	.58	.51	.52	1.00		
.71	.69	.67	.64	.45	.45	.75	.64	1.00	
.78	.68	.76	.69	.46	.46	.79	.61	.75	1.00

(e) Correlation Matrix from Davis (1944).

N = 421

1.00								
.72	1.00							
.41	.34	1.00						
.28	.36	.16	1.00					
.52	.53	.34	.30	1.00				
.71	.71	.43	.36	.64	1.00			
.68	.68	.42	.35	.55	.76	1.00		
.51	.52	.28	.29	.45	.57	.59	1.00	
.68	.68	.41	.36	.55	.76	.68	.58	1.00

Table 3 Analysis of Davis' Data: Initial Factor Loadings and Error Variances

BLR

	1	2	Error Varince
1	0.812843	-0.114657	0.331688
2	0.816658	-0.014656	0.338511
3	0.477504	-0.077121	0.767954
4	0.453904	0.741009	0.246798
5	0.676648	-0.005734	0.546000
6	0.896015	-0.056022	0.200804
7	0.842199	-0.039545	0.295139
8	0.660842	-0.005573	0.566964
9	0.841190	-0.025113	0.297765

BLM

	1	2	Error Varince
1	0.855872	-0.440427	0.073185
2	0.810301	-0.057998	0.339893
3	0.476637	0.000079	0.772769
4	0.405390	0.146887	0.814058
5	0.672354	0.128670	0.531306
6	0.893300	0.124247	0.186432
7	0.834940	0.078102	0.296642
8	0.654167	0.112733	0.559281
9	0.833737	0.076118	0.298954

BHR

	1	2	Error Varince
1	0.810807	-0.122295	0.330455
2	0.813890	-0.010746	0.339987
3	0.476344	-0.087920	0.755056
4	0.439384	0.620345	0.422007
5	0.674253	0.008224	0.541602
6	0.889749	-0.043316	0.212918
7	0.838562	-0.030976	0.299621
8	0.659494	0.012749	0.560605
9	0.837584	-0.013738	0.301959

Table 6 (continued)

(b) Factor Loadings and Factor Scoring Weights
Values at the End of Marginal Estimation of Error Variance.

Factor Loadings

	1	2	3	4
1	0.584022	-0.337871	0.044703	0.088162
2	0.541687	-0.254189	-0.023022	0.048515
3	0.536513	-0.324311	-0.032304	0.060920
4	0.005327	-0.712261	0.005046	-0.000537
5	-0.024393	-0.734924	-0.021617	0.002592
6	-0.018424	-0.730419	-0.016213	0.001663
7	0.085915	-0.651269	0.075602	-0.008608
8	-0.049310	-0.753803	-0.043608	0.005107
9	0.000102	-0.190575	0.598915	0.132620
10	0.060305	-0.291032	0.599291	0.170080
11	0.158430	-0.185615	0.561312	0.043399
12	0.395364	-0.321665	0.524598	0.007841
13	-0.012577	-0.220312	0.118324	0.511894
14	0.010144	-0.169255	0.105527	0.508116
15	0.061011	-0.227214	0.177197	0.504992
16	0.206767	-0.183465	0.260190	0.488820

Sample Mean and Dispersion of Factor Loadings

	1	2	3	4
	0.158805	-0.393011	0.183371	0.160349

	1	2	3	4
1	0.047588	0.017093	-0.003962	-0.009201
2	0.017093	0.050459	0.029291	0.028556
3	-0.003962	0.029291	0.056644	0.004383
4	-0.009201	0.028556	0.004383	0.041575

Table 6 (continued)

(a) (continued)

Factor Scoring Weight

	1	2	3	4
1	0.394413	-0.030042	-0.079415	0.017196
2	0.281880	-0.013292	-0.077284	0.001809
3	0.300054	-0.029260	-0.093438	0.009034
4	-0.068900	-0.220555	-0.044782	-0.049292
5	-0.101037	-0.249806	-0.067369	-0.042865
6	-0.098456	-0.254434	-0.065290	-0.046505
7	-0.000655	-0.150542	0.001760	-0.059018
8	-0.133333	-0.280495	-0.089655	-0.037181
9	-0.090953	-0.007862	0.321248	-0.005793
10	-0.060997	-0.026252	0.312716	0.016743
11	0.028308	-0.000603	0.294939	-0.078495
12	0.225164	-0.021997	0.293564	-0.143138
13	-0.050183	-0.016455	-0.021071	0.291471
14	-0.035100	-0.006594	-0.027347	0.303153
15	-0.018764	-0.012775	-0.000553	0.315139
16	0.064876	0.009493	0.031664	0.288458

Sample Dispersion of Factor Scores

	1	2	3	4
1	0.655222	-0.073111	0.088669	0.023238
2	-0.073111	1.085088	-0.083745	-0.105452
3	0.088669	-0.083745	0.777002	0.089481
4	0.023238	-0.105452	0.089481	0.597317

Table 6 Analysis of Harman's Data

(a) Factor Loadings and Factor Scoring Weights
Values at the End of Estimation of Hyperparameters.

Factor Loadings

	1	2	3	4
1	0.583695	-0.337757	0.044130	0.087959
2	0.541807	-0.254016	-0.022888	0.048518
3	0.536692	-0.324043	-0.032077	0.060933
4	0.005473	-0.712123	0.005168	-0.000551
5	-0.023985	-0.734585	-0.021257	0.002549
6	-0.018178	-0.730202	-0.015998	0.001643
7	0.085502	-0.651552	0.075240	-0.008566
8	-0.048773	-0.753366	-0.043133	0.005051
9	0.001400	-0.190835	0.598659	0.132125
10	0.061218	-0.291066	0.599099	0.169649
11	0.158644	-0.185795	0.561296	0.043424
12	0.394458	-0.321751	0.524847	0.008443
13	-0.011712	-0.220081	0.118846	0.511821
14	0.010753	-0.169247	0.105990	0.508071
15	0.061607	-0.227154	0.177626	0.504945
16	0.206248	-0.183886	0.260088	0.488898

Sample Mean and Dispersion of Factor Loadings

	1	2	3	4
	0.159053	-0.392966	0.183477	0.160307
1	0.047453	0.017109	-0.003968	-0.009156
2	0.017109	0.050408	0.029247	0.028539
3	-0.003968	0.029247	0.056595	0.004396
4	-0.009156	0.028539	0.004396	0.041564

Table 5 (continued)

(g) (continued)

Group 4: Variables 5 and 6.

$\underline{a}_4$ = [.0, .0, ,0, 665442],

C_{A4}

	1	2	3	4
1	0.091414	-0.056729	-0.089281	-0.002489
2	-0.056729	0.038457	0.054028	-0.001978
3	-0.089281	0.054028	0.087783	0.003923
4	-0.002489	-0.001978	0.003923	0.003883

Hyperparameters for the Error Variances

n = 16.058422 s = 3.429397

Table 5 (continued)

(g) Hyperparameters
With Locally Exchangeable Prior with A Priori Zeros

Hyperparameters for the Factor Loadings

Group 1: Variables 1, 3 and 4.

$\underline{a}_1$ = [.724779, .0, .0, .0],

C_{A1}

	1	2	3	4
1	0.004785	0.002168	0.002714	0.001054
2	0.002168	0.179911	0.113805	-0.066294
3	0.002714	0.113805	0.072865	-0.043123
4	0.001054	-0.066294	-0.043123	0.031040

Group 2: Variables 7, 9 and 10.

$\underline{a}_2$ = [.0, -.716819, .0, .0],

C_{A2}

	1	2	3	4
1	0.172095	0.000920	-0.102238	0.060819
2	0.000920	0.002812	-0.004285	0.000267
3	-0.102238	-0.004285	0.070773	-0.037518
4	0.060819	0.000267	-0.037518	0.021920

Group 3: Variables 2 and 8.

$\underline{a}_3$ = [.0, .0, -.664751, .0],

C_{A3}

	1	2	3	4
1	0.165055	-0.141059	0.001925	0.101284
2	-0.141059	0.120626	-0.001293	-0.086814
3	0.001925	-0.001293	0.001697	-0.000033
4	0.101284	-0.086814	-0.000033	0.063032

Table 5 (continued)

(e) Factor Loadings (Varimax Rotation)
MLE by SAS PROC FACTOR with SMC Option

	1	2	3	4	Error Variance
1	0.52646	0.58392	0.21769	0.36209	.20338
2	0.37697	0.35478	0.38362	0.62558	.19352
3	0.42621	0.77377	0.28933	0.24984	.07349
4	0.39202	0.61580	0.36376	0.28062	.25605
5	0.15159	0.22804	0.73929	0.27133	.30485
6	0.23203	0.17051	0.70087	0.17277	.39602
7	0.81020	0.29513	0.22335	0.16403	.17969
8	0.29934	0.26396	0.38636	0.69531	.20799
9	0.67360	0.27751	0.25067	0.38481	.25834
10	0.60873	0.41570	0.24666	0.28149	.19890

(f) Factor Loadings
(Oblique Rotation from Ree, el.al (1981))

	1	2	3	4
1	.54	.27	.26	-.04
2	.21	.15	.59	.14
3	.70	.16	.13	.08
4	.62	.12	.15	.57
5	.13	.08	.19	.57
6	.07	.20	.10	.56
7	.23	.68	.04	.01
8	.10	.12	.62	.17
9	.13	.58	.29	.00
10	.33	.56	.14	.02

Table 5 (continued)

(c) Factor Loadings (Varimax Rotation)
With Locally Exchangeable Prior

	1	2	3	4
1	0.5440	0.2128	-0.3580	-0.5906
2	0.3905	0.3750	-0.6304	-0.3650
3	0.4487	0.2879	-0.2563	-0.7573
4	0.4006	0.3459	-0.2893	-0.6332
5	0.1639	0.6747	-0.3099	-0.2516
6	0.2259	0.7870	-0.1666	-0.1626
7	0.8133	0.2143	-0.1720	-0.2960
8	0.3180	0.3782	-0.7160	-0.2700
9	0.6899	0.2480	-0.3850	-0.2776
10	0.7143	0.2449	-0.2797	-0.4131

(d) Factor Loadings (Varimax Rotation)
With Globally Exchangeable Prior

	1	2	3	4
1	0.5311	0.2211	-0.3624	-0.5924
2	0.3720	0.3724	-0.6205	-0.3583
3	0.4328	0.2894	-0.2534	-0.7586
4	0.3913	0.3621	-0.2788	-0.6252
5	0.1521	0.7155	-0.2794	-0.2350
6	0.2347	0.7190	-0.1691	-0.1677
7	0.8105	0.2242	-0.1670	-0.2968
8	0.3059	0.3921	-0.6988	-0.2665
9	0.6746	0.2508	-0.3850	-0.2788
10	0.6936	0.2382	-0.2799	-0.4174

Table 5 Analysis of ASVAB Form 8A

(a) Factor Loadings (without Rotation)
With Locally Exchangeable Prior with A Priori Zeros

	1	2	3	4
1	0.6581	−0.5029	−0.3158	0.0948
2	0.4516	−0.3784	−0.6236	0.2511
3	0.8200	−0.3936	−0.2161	0.1678
4	0.6998	−0.3598	−0.2662	0.2375
5	0.3398	−0.1627	−0.3523	0.6040
6	0.2594	−0.2249	−0.2263	0.7275
7	0.3784	−0.7902	−0.1407	0.1401
8	0.3574	−0.3147	−0.7048	0.2514
9	0.3641	−0.6724	−0.3596	0.1529
10	0.4950	−0.6877	−0.2474	0.1507

(b) Factor Loadings (Varimax Rotation)
With Locally Exchangeable Prior with A Priori Zeros

	1	2	3	4
1	0.5852	−0.5346	−0.3532	0.2043
2	0.3562	−0.3805	−0.6258	0.3684
3	0.7532	−0.4394	−0.2514	0.2802
4	0.6271	−0.3936	−0.2857	0.3415
5	0.2457	−0.1595	−0.3040	0.6727
6	0.1576	−0.2175	−0.1627	0.7744
7	0.2928	−0.8062	−0.1682	0.2076
8	0.2627	−0.3073	−0.6992	0.3665
9	0.2728	−0.6794	−0.3793	0.2399
10	0.4085	−0.7083	−0.2759	0.2393

Table 4 (continued)

(d) Final Value of $F^{*\prime}F^{*}$.

Solution with Low Error Varinace on Variable #4.

	1	2
1	0.937479	0.005313
2	0.005313	0.607890

Solution with High Error Varinace on Variable #4.

	1	2
1	0.940400	-0.019373
2	-0.019373	0.364831

Table 4 (continued)

(c) Final Conditional Expectation of Factor Score Weights.

Solution with Low Error Varinace on Variable #4.

	1	2
1	0.155969	-0.153019
2	0.151353	-0.027197
3	0.040342	-0.046109
4	0.075110	0.833996
5	0.078564	-0.006782
6	0.266449	-0.116043
7	0.177608	-0.062457
8	0.074163	-0.005465
9	0.175636	-0.041495

Solution with High Error Varinace on Variable #4.

	1	2
1	0.185876	-0.762509
2	0.144439	-0.246647
3	0.037639	0.009005
4	0.034176	0.118204
5	0.083909	0.198785
6	0.276851	0.363198
7	0.170171	0.126135
8	0.075363	0.136785
9	0.168515	0.121119

Table 4 Analysis of Davis' Data: Results

(a) Final Conditional Expectation of Error Varinaces

Solution with Low Error Varinace on Variable #4.

0.328924 0.336681 0.749110 0.322409 0.536633 0.209846
0.296318 0.555788 0.298808

Solution with High Error Varinace on Variable #4.

0.232446 0.316048 0.754362 0.793634 0.513588 0.198474
0.297366 0.546577 0.299606

(b) Final Conditional Expectation of Factor Loadings.

Solution with Low Error Varinace on Variable #4.

	1	2
1	0.810576	-0.117377
2	0.813995	-0.012322
3	0.476424	-0.081633
4	0.446113	0.691789
5	0.674320	-0.000144
6	0.890054	-0.050043
7	0.838720	-0.035834
8	0.659376	0.001189
9	0.837695	-0.020270

Solution with High Error Varinace on Variable #4.

	1	2
1	0.823812	-0.304174
2	0.813901	-0.147551
3	0.477665	-0.003874
4	0.411161	0.135154
5	0.677542	0.140070
6	0.893923	0.086225
7	0.838166	0.033489
8	0.659422	0.097594
9	0.836848	0.031608

Table 3 (continued)

SX#7

	1	2	Error Varince
1	0.680000	0.443260	0.341120
2	0.680000	0.447630	0.337230
3	0.420000	0.222940	0.773900
4	0.350000	0.212910	0.832170
5	0.550000	0.393430	0.542710
6	0.760000	0.464740	0.206240
7	1.000000	0.000000	0.000000
8	0.590000	0.287690	0.569130
9	0.680000	0.507700	0.279840

SX#9

	1	2	Error Varince
1	0.680000	0.442940	0.341400
2	0.680000	0.446050	0.338640
3	0.410000	0.243240	0.772730
4	0.360000	0.192510	0.889333
5	0.550000	0.392760	0.543240
6	0.760000	0.464030	0.207070
7	0.680000	0.510450	0.277040
8	0.580000	0.307850	0.568830
9	1.000000	0.000000	0.000000

Table 3 (continued)

SSMC, SBLM, SBHR, SBHM, SBMR, and SBMH

	1	2	Error Varince
1	1.000000	0.000000	0.000000
2	0.720000	0.369970	0.344720
3	0.410000	0.242340	0.773170
4	0.280000	0.326190	0.815200
5	0.520000	0.443850	0.532600
6	0.710000	0.556080	0.186680
7	0.680000	0.491090	0.296430
8	0.510000	0.424890	0.559370
9	0.680000	0.488650	0.298820

SBLR

	1	2	Error Varince
1	0.280000	0.766790	0.333630
2	0.360000	0.728960	0.339010
3	0.160000	0.453180	0.769030
4	1.000000	0.000000	0.000000
5	0.300000	0.603130	0.546230
6	0.360000	0.818510	0.200450
7	0.350000	0.763130	0.295130
8	0.290000	0.590430	0.567290
9	0.360000	0.756610	0.297930

SX#2

	1	2	Error Varince
1	0.720000	0.365890	0.347720
2	1.000000	0.000000	0.000000
3	0.340000	0.357170	0.756830
4	0.360000	0.195780	0.832070
5	0.530000	0.427760	0.536120
6	0.710000	0.558710	0.183750
7	0.680000	0.491730	0.295800
8	0.520000	0.406140	0.564650
9	0.680000	0.488240	0.299220

Table 3 (continued)

BHM

	1	2	Error Varince
1	0.811119	-0.121935	0.330069
2	0.813922	-0.009310	0.339990
3	0.476510	-0.088143	0.754991
4	0.437326	0.610792	0.435266
5	0.674248	0.010934	0.541635
6	0.889839	-0.040549	0.212984
7	0.838630	-0.028670	0.299661
8	0.659503	0.016008	0.560582
9	0.837628	-0.011161	0.301968

BMR

	1	2	Error Varince
1	0.836022	-0.382014	0.155784
2	0.811075	-0.103317	0.333018
3	0.476852	-0.012453	0.776130
4	0.408311	0.139907	0.817584
5	0.675165	0.123325	0.531444
6	0.896054	0.102585	0.187412
7	0.836854	0.049606	0.298595
8	0.656418	0.097122	0.562332
9	0.835626	0.047953	0.300820

BMM

	1	2	Error Varince
1	0.836910	-0.380007	0.155828
2	0.811316	-0.101433	0.333012
3	0.476880	-0.011333	0.776130
4	0.407981	0.140862	0.817585
5	0.674874	0.124915	0.531442
6	0.895811	0.104686	0.187411
7	0.836735	0.051562	0.298595
8	0.656188	0.098659	0.562332
9	0.835511	0.049907	0.300820

Table 6 (continued)

(b) (continued)

Factor Scoring Weight

	1	2	3	4
1	0.396668	-0.030005	-0.079840	0.017604
2	0.282561	-0.013289	-0.077726	0.002023
3	0.300887	-0.029298	-0.094023	0.009303
4	-0.069127	-0.220696	-0.044808	-0.049597
5	-0.101663	-0.250306	-0.067666	-0.043101
6	-0.098912	-0.254818	-0.065462	-0.046813
7	-0.000359	-0.150332	0.002056	-0.059417
8	-0.134299	-0.281308	-0.090174	-0.037372
9	-0.092510	-0.007878	0.322904	-0.005700
10	-0.062101	-0.026296	0.313909	0.016912
11	0.028051	-0.000590	0.296059	-0.079150
12	0.226944	-0.022005	0.294864	-0.144876
13	-0.050628	-0.016448	-0.021576	0.292930
14	-0.035381	-0.006523	-0.027898	0.304816
15	-0.019035	-0.012715	-0.001049	0.316910
16	0.065508	0.009780	0.031427	0.289864

Sample Dispersion of Factor Scores

	1	2	3	4
1	0.661182	-0.072420	0.088229	0.022859
2	-0.072420	1.087983	-0.083686	-0.105571
3	0.088229	-0.083686	0.783112	0.088769
4	0.022859	-0.105571	0.088769	0.603667

Table 6 (continued)

(c) Factor Loadings and Factor Scoring Weights
Values at the End of Joint Estimation of Factor Scores and Factor Loadings.

Factor Loadings

	1	2	3	4
1	0.612345	-0.599289	0.087146	0.158986
2	0.564085	-0.457232	0.010543	0.103659
3	0.558300	-0.540573	0.000103	0.118598
4	-0.114895	-0.803456	-0.100861	0.011532
5	-0.147203	-0.828065	-0.129800	0.014875
6	-0.137196	-0.820510	-0.120796	0.013547
7	-0.005696	-0.720830	-0.004974	0.000636
8	-0.175978	-0.849867	-0.155236	0.017726
9	-0.331131	-0.159671	0.672917	0.292350
10	-0.188156	-0.314258	0.661643	0.322302
11	-0.018781	-0.272320	0.616332	0.201005
12	0.429215	-0.554887	0.551572	0.152400
13	-0.184234	-0.112750	-0.071843	0.525962
14	-0.158000	-0.059276	-0.083435	0.521776
15	-0.103943	-0.162369	0.016787	0.519876
16	0.129169	-0.209889	0.217741	0.498008

Sample Mean and Dispersion of Factor Loadings

	1	2	3	4
	0.045494	-0.466578	0.135490	0.217077
1	0.092029	-0.012986	-0.002124	-0.013083
2	-0.012986	0.076038	0.034554	0.049943
3	-0.002124	0.034554	0.088389	0.011916
4	-0.013083	0.049943	0.011916	0.038570

Table 6 (continued)

(c) (continued)

Factor Scoring Weight

	1	2	3	4
1	0.321475	-0.076134	-0.011602	0.073320
2	0.221562	-0.041760	-0.028399	0.043020
3	0.237909	-0.056945	-0.037947	0.051954
4	-0.101998	-0.180280	-0.062740	-0.053232
5	-0.132437	-0.200977	-0.086976	-0.050169
6	-0.129012	-0.204547	-0.083143	-0.054189
7	-0.024690	-0.126345	-0.002552	-0.059505
8	-0.164554	-0.223292	-0.112158	-0.047452
9	-0.190523	-0.013054	0.289455	0.047816
10	-0.123135	-0.033973	0.281129	0.066094
11	-0.032573	-0.030558	0.267324	-0.004878
12	0.229642	-0.083671	0.274954	-0.040951
13	-0.059283	0.014240	-0.105013	0.288855
14	-0.049334	0.022418	-0.113743	0.303809
15	-0.035622	0.010478	-0.082566	0.308744
16	0.061710	0.007897	0.001744	0.263956

Sample Dispersion of Factor Scores

	1	2	3	4
1	0.526156	0.118747	0.035452	-0.005656
2	0.118747	0.841797	-0.030994	-0.197943
3	0.035452	-0.030994	0.620489	0.091274
4	-0.005656	-0.197943	0.091274	0.707033

Table 6 (continued)

(d) Factor Loadings and Factor Scoring Weights
Values at the End of Marginal Estimation of Factor Loadings.

Factor Loadings

	1	2	3	4
1	0.590076	-0.561924	0.052008	0.139681
2	0.549209	-0.429081	-0.012859	0.090081
3	0.543649	-0.505775	-0.022812	0.103710
4	-0.079764	-0.776750	-0.069955	0.008106
5	-0.106069	-0.796811	-0.093605	0.010848
6	-0.097399	-0.790265	-0.085782	0.009661
7	0.010937	-0.708142	0.009638	-0.000935
8	-0.129254	-0.814379	-0.114118	0.013140
9	-0.208041	-0.187953	0.647443	0.244572
10	-0.103776	-0.318461	0.641745	0.278818
11	0.013599	-0.268345	0.607520	0.180310
12	0.332875	-0.493583	0.564576	0.161941
13	-0.124651	-0.136912	-0.013462	0.520344
14	-0.101330	-0.088096	-0.024507	0.516577
15	-0.062553	-0.177377	0.056320	0.515738
16	0.116852	-0.210490	0.208957	0.498821

Sample Mean and Dispersion of Factor Loadings

	1	2	3	4
	0.071523	-0.454021	0.146944	0.205713
1	0.069742	-0.005311	-0.006141	-0.010388
2	-0.005311	0.064975	0.032524	0.044741
3	-0.006141	0.032524	0.078656	0.010502
4	-0.010388	0.044741	0.010502	0.038153

Table 6 (continued)

(d) (continued)

Factor Scoring Weight

	1	2	3	4
1	0.310975	−0.071712	−0.025250	0.068637
2	0.216437	−0.039193	−0.034988	0.040520
3	0.232531	−0.053251	−0.044705	0.048807
4	−0.079177	−0.173407	−0.043572	−0.060132
5	−0.103658	−0.192315	−0.062792	−0.058892
6	−0.100379	−0.195927	−0.059090	−0.062856
7	−0.016103	−0.123747	0.004636	−0.062054
8	−0.129295	−0.212685	−0.082522	−0.058155
9	−0.128161	−0.018218	0.282117	0.020503
10	−0.078863	−0.035845	0.276962	0.040581
11	−0.016123	−0.030625	0.266015	−0.016500
12	0.171872	−0.073546	0.284514	−0.033907
13	−0.038000	0.012078	−0.083439	0.276675
14	−0.028535	0.019496	−0.091102	0.290919
15	−0.018683	0.009014	−0.065872	0.299253
16	0.056413	0.007544	−0.001865	0.265621

Sample Dispersion of Factor Scores

	1	2	3	4
1	0.526156	0.118767	0.035453	−0.005663
2	0.118767	0.841779	−0.030984	−0.197950
3	0.035453	−0.030984	0.620473	0.091274
4	−0.005663	−0.197950	0.091274	0.707048

Table 6 (continued)

(e) Factor Loadings and Factor Scoring Weights
Values at the End of Marginal Estimation of Factor Scores.

Factor Loadings

	1	2	3	4
1	0.612883	-0.620351	0.087765	0.163808
2	0.564815	-0.472804	0.011541	0.107391
3	0.559069	-0.557269	0.001153	0.122592
4	-0.120888	-0.808003	-0.106144	0.012138
5	-0.153182	-0.832600	-0.135066	0.015476
6	-0.143460	-0.825260	-0.126312	0.014180
7	-0.012304	-0.725842	-0.010794	0.001311
8	-0.181844	-0.854317	-0.160402	0.018312
9	-0.342213	-0.162270	0.675906	0.300225
10	-0.191989	-0.322731	0.663786	0.330311
11	-0.024473	-0.283294	0.619279	0.211781
12	0.436680	-0.578528	0.553292	0.164836
13	-0.194428	-0.104943	-0.083961	0.526721
14	-0.169842	-0.050546	-0.097312	0.522669
15	-0.114489	-0.158021	0.006412	0.520785
16	0.122458	-0.214506	0.215417	0.498839

Sample Mean and Dispersion of Factor Loadings

	1	2	3	4
	0.040425	-0.473205	0.132160	0.220711
1	0.094559	-0.015765	-0.001086	-0.013216
2	-0.015765	0.078125	0.033405	0.051106
3	-0.001086	0.033405	0.090609	0.012573
4	-0.013216	0.051106	0.012573	0.038553

Table 6 (continued)

(e) (continued)

Factor Scoring Weight

	1	2	3	4
1	0.310729	-0.076873	-0.014801	0.077513
2	0.214367	-0.042087	-0.030034	0.045990
3	0.230309	-0.057105	-0.039526	0.054979
4	-0.101932	-0.176757	-0.062581	-0.055287
5	-0.131321	-0.196949	-0.086196	-0.052904
6	-0.128394	-0.200537	-0.082699	-0.056773
7	-0.027509	-0.124206	-0.004153	-0.059769
8	-0.162279	-0.218703	-0.110677	-0.050922
9	-0.190825	-0.012891	0.286579	0.047963
10	-0.122128	-0.033905	0.277122	0.066889
11	-0.036273	-0.031106	0.262660	-0.000217
12	0.223221	-0.085000	0.266164	-0.030328
13	-0.056747	0.016223	-0.107366	0.286299
14	-0.047645	0.024393	-0.116936	0.301758
15	-0.034660	0.012265	-0.085416	0.306006
16	0.059241	0.008256	-0.000125	0.261205

Sample Dispersion of Factor Scores

	1	2	3	4
1	0.502935	0.126244	0.025704	-0.003952
2	0.126244	0.813467	-0.021044	-0.192335
3	0.025704	-0.021044	0.597943	0.087372
4	-0.003952	-0.192335	0.087372	0.705390

Table 6 (continued)

(f) Factor Loadings and Factor Scoring Weights
Error Variances at the End of Hyperparameter Estimation.

0.463089	0.625698	0.573485	0.383050	0.355957	0.346092
0.481711	0.330248	0.491035	0.475497	0.493230	0.411494
0.644330	0.620561	0.565170	0.562536		

MEAN AND VARIANCE OF ERROR VARIANCES = 0.488949 0.009997

(g) Factor Loadings and Factor Scoring Weights
Error Variances at the End of Marginal Estimate of Error Variance.

0.455959	0.617735	0.565945	0.378151	0.351078	0.341424
0.476000	0.325488	0.483854	0.468998	0.486393	0.405130
0.635949	0.612180	0.557435	0.554944		

MEAN AND VARIANCE OF ERROR VARIANCES = 0.482291 0.009758

Table 6 (continued)

(h) Factor Loadings and Factor Scoring Weights Hyperparameters

Hyperparameters of the Error Variances

$n = 35.509313$ $s = 16.601690$

Hyperparameters of the Factor Loadings

Group 1: Variables # 1, 2, and 3.

$\underline{a}_1 = [\ .554064,\ .0\ ,.0\ .0\ ]$

C_{A1}

	1	2	3	4
1	0.000901	−0.000338	0.001448	0.000530
2	−0.000338	0.097439	0.000604	−0.021174
3	0.001448	0.000604	0.002367	0.000611
4	0.000530	−0.021174	0.000611	0.004843

Group 2: Variables # 4, 5, 6, 7 and 9.

$\underline{a}_2 = [\ .0,\ -.716366,\ .0,\ .0\ ]$

C_{A2}

	1	2	3	4
1	0.002903	0.002197	0.002550	−0.000292
2	0.002197	0.001671	0.001935	−0.000221
3	0.002550	0.001935	0.002258	−0.000258
4	−0.000292	−0.000221	−0.000258	0.000039

Table 6 (continued)

(h) (continued)

Group 3: Variables # 9, 10, 11, and 12.

$\underline{a}_3$ = [.0, .0, .570975, .0]

C_{A3}

	1	2	3	4
1	0.048505	−0.044281	−0.004950	0.004370
2	−0.044281	0.066629	0.000686	−0.022250
3	−0.004950	0.000686	0.001085	0.002238
4	0.004370	−0.022250	0.002238	0.013150

Group 4: Variables # 13, 14, 15, and 16.

$\underline{a}_4$ = [.0, .0, .0, .503434]

C_{A4}

	1	2	3	4
1	0.013770	−0.012277	0.017359	−0.000967
2	−0.012277	0.043087	−0.034108	−0.000208
3	0.017359	−0.034108	0.032701	−0.000599
4	−0.000967	−0.000208	−0.000599	0.000115

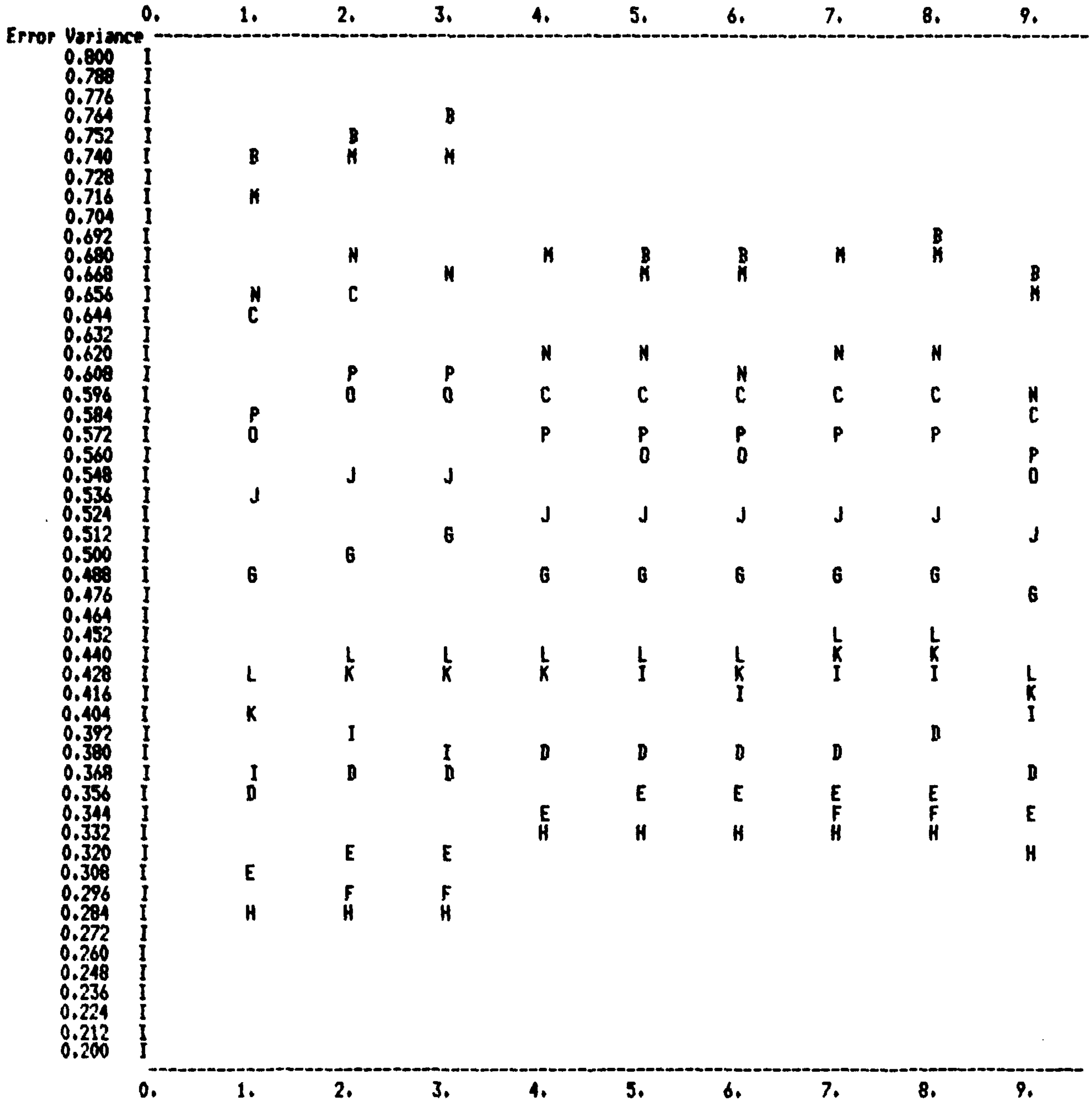


Figure. 1 Various Estimates of Error Variances of the Harman Data

N = 145

	D.F.	LAMBDA	MEAN	VARIANCE
1: M.L.E. BY SAS PROC FACTOR			0.477034	0.022715
2: MARGINAL ESTIMATE OF E-VAR BY E-AF			0.495399	0.023301
3: MARGINAL ESTIMATE OF E-VAR BY E-FA			0.494445	0.023843
4: LOADINGS BY FORMULA, E-VAR BY LOG-NORMAL	26.826477	0.473865	0.484629	0.013195
5: LOADINGS BY E-AF, ERR-VAR BY LOG-NORMAL	29.735748	0.474122	0.484497	0.012549
6: LOADINGS BY E-FA, ERR-VAR BY LOG-NORMAL	29.842918	0.473667	0.484009	0.012542
7: LOADINGS BY E-AF, ERR-VAR BY EM AFTER E-AF	30.446888	0.463148	0.489141	0.012720
8: LOADINGS BY E-FA, ERR-VAR BY EM AFTER E-FA	30.610523	0.462792	0.488639	0.012693
9: LOADINGS BY E-AF, ERR-VAR BY EM AFTER E-F	28.854729	0.444613	0.471290	0.012644

LAMBDA = s x n

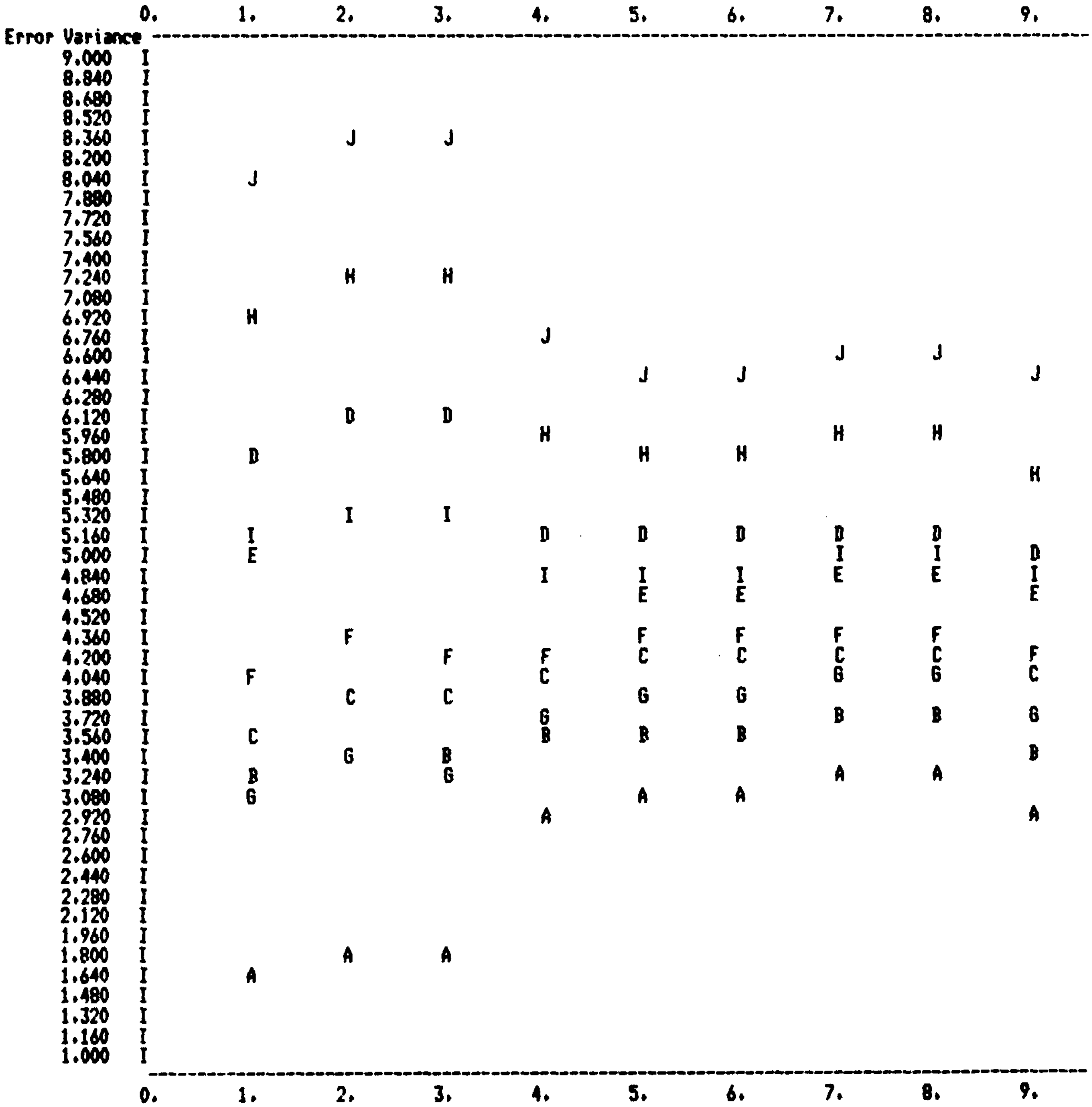
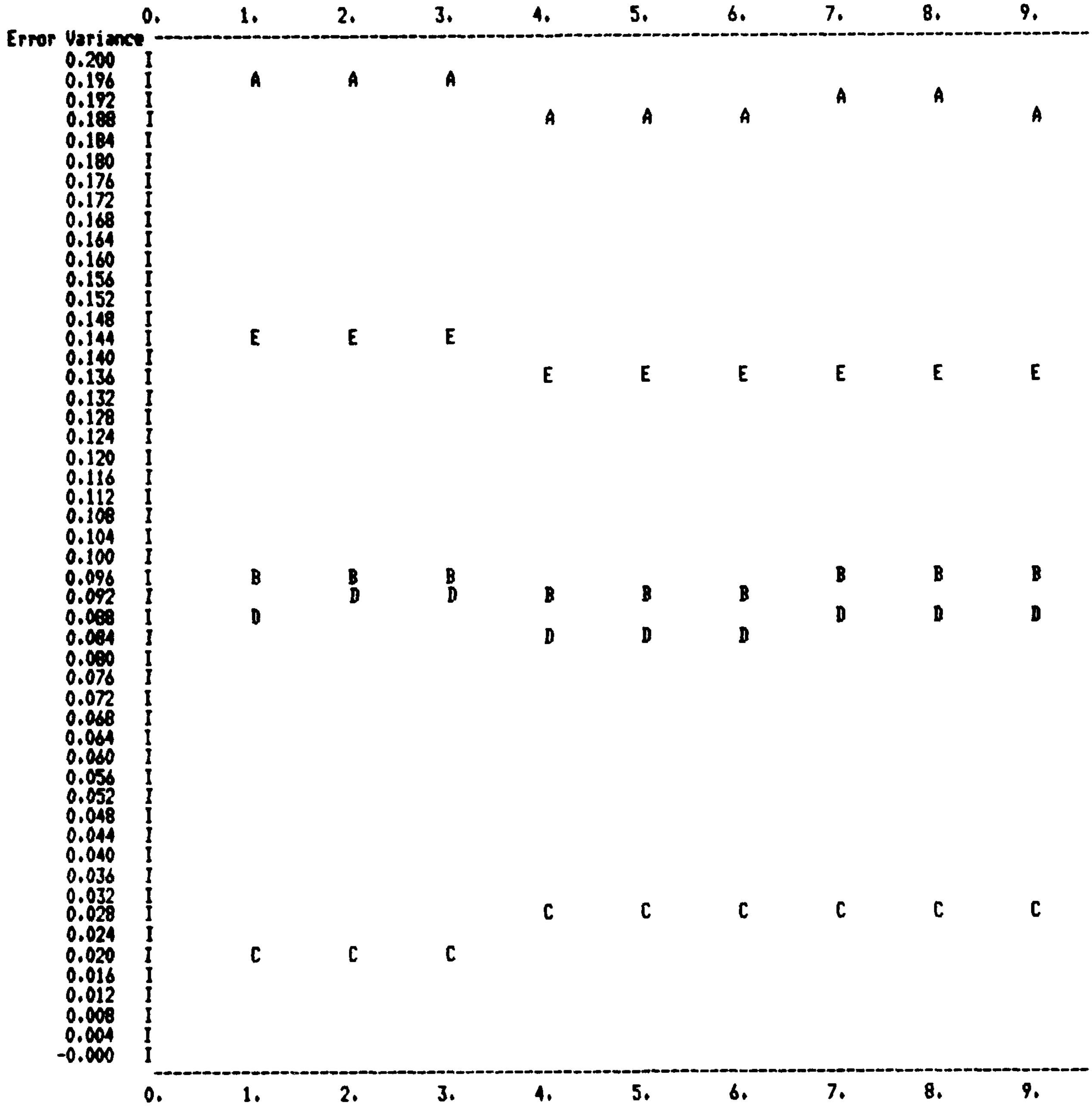


Figure. 2 Various Estimates of Error Variances of the Francis Data

N = 50

	D.F.	LAMBDA	MEAN	VARIANCE
1: M.L.E. BY SAS PROC FACTOR			4.575092	3.291272
2: MARGINAL ESTIMATE OF E-VAR BY E-AF			4.806640	3.549344
3: MARGINAL ESTIMATE OF E-VAR BY E-FA			4.802694	3.565243
4: LOADINGS BY FORMULA, E-VAR BY LOG-NORMAL	20.144287	4.458605	4.521993	1.203936
5: LOADINGS BY E-AF, ERR-VAR BY LOG-NORMAL	26.009495	4.468840	4.519325	0.977160
6: LOADINGS BY E-FA, ERR-VAR BY LOG-NORMAL	25.998222	4.467546	4.518114	0.977998
7: LOADINGS BY E-AF, ERR-VAR BY EM AFTER E-AF	29.433685	4.442304	4.638591	0.923927
8: LOADINGS BY E-FA, ERR-VAR BY EM AFTER E-FA	29.431026	4.440956	4.637250	0.924261
9: LOADINGS BY E-AF, ERR-VAR BY EM AFTER E-F	25.810238	4.187857	4.409835	0.992775

LAMBDA = s x n



Fisure. 3 Various Estimates of Error Variances of the Mardia, et., al. Data

N = 100

	D.F.	LAMBDA	MEAN	VARIANCE
1: M.L.E. BY SAS PROC FACTOR			0.106888	0.003425
2: MARGINAL ESTIMATE OF E-VAR BY E-AF			0.107966	0.003465
3: MARGINAL ESTIMATE OF E-VAR BY E-FA			0.108053	0.003517
4: LOADINGS BY FORMULA, E-VAR BY LOG-NORMAL	3.182747	0.086875	0.103969	0.002884
5: LOADINGS BY E-AF, ERR-VAR BY LOG-NORMAL	4.215380	0.087360	0.103681	0.002780
6: LOADINGS BY E-FA, ERR-VAR BY LOG-NORMAL	4.131446	0.087010	0.103574	0.002800
7: LOADINGS BY E-AF, ERR-VAR BY EM AFTER E-AF	3.838849	0.066296	0.105307	0.002966
8: LOADINGS BY E-FA, ERR-VAR BY EM AFTER E-FA	3.743914	0.065507	0.105242	0.002991
9: LOADINGS BY E-AF, ERR-VAR BY EM AFTER E-F	3.757377	0.064932	0.104149	0.002926

LAMBDA = s x n

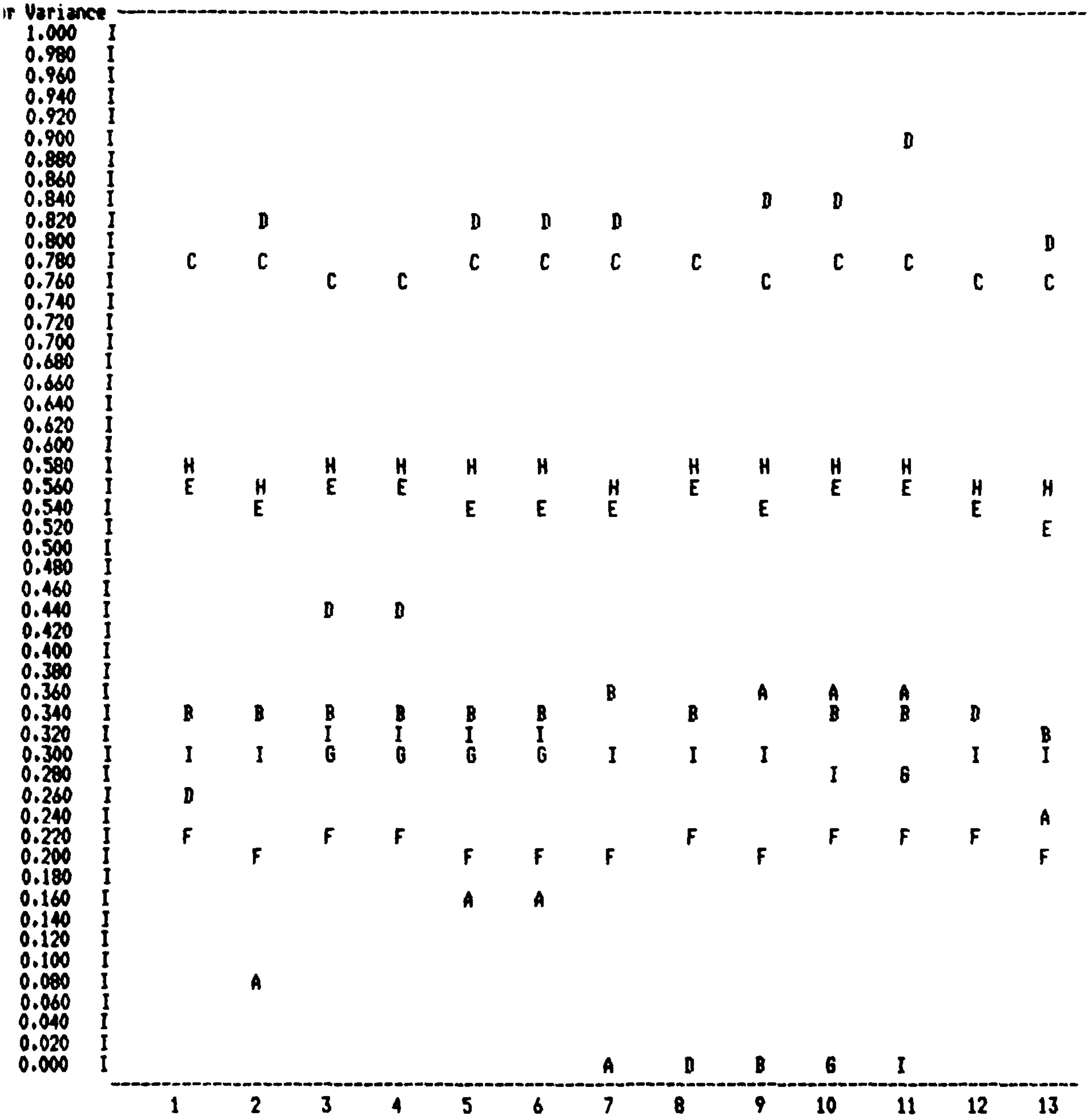


Figure. 4 Various Estimates of Error Variances of the Davis Data

GRAPH # 1 = BLR
GRAPH # 2 = BLM
GRAPH # 3 = BHR
GRAPH # 4 = BHM
GRAPH # 5 = BMR
GRAPH # 6 = BMM
GRAPH # 7 = SSMC SBLM SBHR SBHM SBMR SBMM
GRAPH # 8 = SBLR
GRAPH # 9 = SX#2
GRAPH # 10 = SX#7
GRAPH # 11 = SX#9
GRAPH # 12 = BKBLR BKBHR BKBHM BKSBLR
GRAPH # 13 = BKBLM BKBMR BMBMM BKSSMC BKSBLM BKSBHR BKSBHM BKSBMR BKSBMM BKSX#2 BKSX#7 BKSX#9

University of Iowa/Novick 8 March 1985

Distribution List

Personnel Analysis Division
AF/MPXA
5C360, The Pentagon
Washington, DC 20330

Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

Air Force Office
of Scientific Research
Life Sciences Directorate
Bolling Air Force Base
Washington, DC 20332

Dr. James Algina
University of Florida
Gainesville, FL 32605

Technical Director
Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Special Assistant for Projects
OASN(M&RA)
5D800, The Pentagon
Washington, DC 20350

Dr. R. Darrell Bock
University of Chicago
Department of Education
Chicago, IL 60637

Dr. Nick Bond
Office of Naval Research
Liaison Office, Far East
APO San Francisco, CA 96503

Dr. Robert Brennan
American College Testing
Programs
P. O. Box 168
Iowa City, IA 52243

Col. Roger Campbell
AF/MPXOA
Pentagon, Room 4E195
Washington, DC 20330

Dr. James Carlson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52243

Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

Director
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Scientific Advisor
to the DCNO (MPT)
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Chief of Naval Education
and Training
Liason Office
AFHRL
Operations Training Division
Williams AFB, AZ 85224

Office of the Chief
of Naval Operations
Research Development
& Studies Branch
NAVOP 01B7
Washington, DC 20350

Dr. Stanley Collyer
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217

CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217

Mr. Timothy Davey
University of Illinois
Educational Psychology
Urbana, IL 61801

Dr. Dattprasad Divgi
Syracuse University
Department of Psychology
Syracuse, NY 13210

Dr. Fritz Drasgow
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
(12 Copies)

Dr. Stephen Dunbar
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

Dr. John M. Eddins
University of Illinois
252 Engineering Research
Laboratory
103 South Mathews Street
Urbana, IL 61801

Dr. Richard Elster
Deputy Assistant Secretary
of the Navy (Manpower)
Washington, DC 20350

ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

Dr. Benjamin A. Fairbank
Performance Metrics, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

Dr. Marshall J. Farr
2520 North Vernon Street
Arlington, VA 22207

Dr. Leonard Feldt
Lindquist Center
for Measurment
University of Iowa
Iowa City, IA 52242

Dr. Myron Fischl
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Dr. Ron Hambleton
School of Education
University of Massachusetts
Amherst, MA 01002

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 62
San Diego, CA 92152

Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 90010

Mr. Dick Hoshaw
NAVOP-135
Arlington Annex
Room 2834
Washington, DC 20350

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Douglas H. Jones
Advanced Statistical
Technologies Corporation
10 Trafalgar Court
Lawrenceville, NJ 08148

Dr. William Koch
University of Texas-Austin
Measurement and Evaluation
Center
Austin, TX 78703

Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Anita Lancaster
Accession Policy
OASD/MI&L/MP&FM/AP
Pentagon
Washington, DC 20301

Dr. Marcy Lansman
University of North Carolina
The L. L. Thurstone Lab.
Davie Hall 013A
Chapel Hill, NC 27514

Dr. Jerry Lehnus
OASD (M&RA)
Washington, DC 20301

Dr. Thomas Leonard
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, WI 53705

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
The NETHERLANDS

Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

Dr. Robert Lockman
Center for Naval Analysis
200 North Beauregard St.
Alexandria, VA 22311

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

Dr. Clessen Martin
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

Dr. James McBride
Psychological Corporation
c/o Harcourt, Brace,
Javanovich Inc.
1250 West 6th Street
San Diego, CA 92101

Dr. Clarence McCormick
HQ, MEPCOM
MEPCT-P
2500 Green Bay Road
North Chicago, IL 60064

Mr. Robert McKinley
University of Toledo
Dept of Educational Psychology
Toledo, OH 43606

Dr. Barbara Means
Human Resources
Research Organization
300 North Washington
Alexandria, VA 22314

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Dr. Karen Mitchell
Army Research Institute
5001 Eisenhower Blvd
Alexandria, VA 22333

Ms. Kathleen Moreno
Navy Personnel R&D Center
Code 62
San Diego, CA 92152

Headquarters, Marine Corps
Code MPI-20
Washington, DC 20380

Lt. Col. Jim Murphy
HQ, Marine Corps
Code MRRP
Washington, DC 20380

Director
Research & Analysis Division
Navy Recruiting Command (Code 22)
4015 Wilson Blvd.
Arlington, VA 22203

Program Manager for Manpower,
Personnel, and Training
NAVMAT 0722
Arlington, VA 22217

Assistant for MPT Research,
Development and Studies
NAVOP 01B7
Washington, DC 20370

Head, Military Compensation
Policy Branch
NAVOP 134
Washington, DC 20370

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73069

Director, Manpower and Personnel
Laboratory
NPRDC (Code 06)
San Diego, CA 92152

Library
Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

Dr. James Olson
WICAT, Inc.
1875 South State Street
Orem, UT 84057

Group Psychology Program
Code 442GP
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217-5000

Office of Naval Research
Code 442PT
800 N. Quincy Street
Arlington, VA 22217
(6 Copies)

Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

Commanding Officer
Army Research Institute
ATTN: PERI-BR (Dr. J. Orasanu)
5001 Eisenhower Avenue
Alexandria, VA 22333

Military Assistant for Training
and Personnel Technology
OUSD(R&E)
Room 3D129, The Pentagon
Washington, DC 20301

Dr. Randolph Park
AFHRL/MOAN
Brooks AFB, TX 78235

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Cirle, NW
Washington, DC 20036

Daira Paulson
Code 52 - Training Systems
Navy Personnel R&D Center
San Diego, CA 92152

Dr. James Paulson
Dept. of Psychology
Portland State Unive3rsity
P.O. Box 751
Portland, OR 97207

Dr. Sam Pearson
CACI, Inc. FED.
Department 5921
1815 North Fort Myers Dr.
Arlington, VA 22209

Dr. James W. Pellegrino
University of California,
Santa Barbara
Dept. of Psychology
Santa Barabara, CA 93106

Administrative Sciences Dept.
Naval Postgraduate School
Monterey, CA 93940

Department of Operations Research
Naval Postgraduate School
Monterey, CA 93940

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
Knoxville, TN 37916

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152

Dr. Robert Sasmor
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Dr. Mary Schratz
Navy Personnel R&D Center
San Diego, CA 92152

Dr. W. Steve Sellman
OASD(MRA&L)
2B269 The Pentagon
Washington, DC 20301

Dr. Joyce Shields
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujusawa 251
JAPAN

Dr. William Sims
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

Dr. H. Wallace Sinaiko
Manpower Research
and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

Dr. Alfred F. Smode
Senior Scientist
Code 7B
Naval Training Equipment Center
Orlando, FL 32813

Dr. Richard Snow
Liaison Scientist
Office of Naval Research
Branch Office, London
Box 39
FPO New York, NY 09510

Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. Peter Stoloff
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

Dr. William Stout
University of Illinois
Department of Mathematics
Urbana, IL 61801

Maj. Bill Strickland
AF/MPXOA
4E168 Pentagon
Washington, DC 20330

Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Brad Sympson
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

Mr. Gary Thomasson
University of Illinois
Educational Psychology
Champaign, IL 61820

Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E. Street, NW
Washington, DC 20415

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 310
St. Paul, MN 55114

Dr. Frank Vicino
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08540

Dr. Ming-Mei Wang
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Mr. Thomas A. Warm
Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Dr. Brian Waters
HumRRO
300 North Washington
Alexandria, VA 22314

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

Dr. Ronald Weitzman
Naval Postgraduate School
Administrative Sciences Dept.
Monterey, CA 93940

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90007

German Military Representative
ATTN: Wolfgang Wildegrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington, DC 20016

Dr. Bruce Williams
Department of Educational
Psychology
University of Illinois
Urbana, IL 61801

Dr. Hilda Wing
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Dr. Martin F. Wiskoff
Navy Personnel R & D Center
San Diego, CA 92152

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152

Dr. George Wong
Biostatistics Laboratory
Memorial Sloan-Kettering
Cancer Center
1275 York Avenue
New York, NY 10021

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

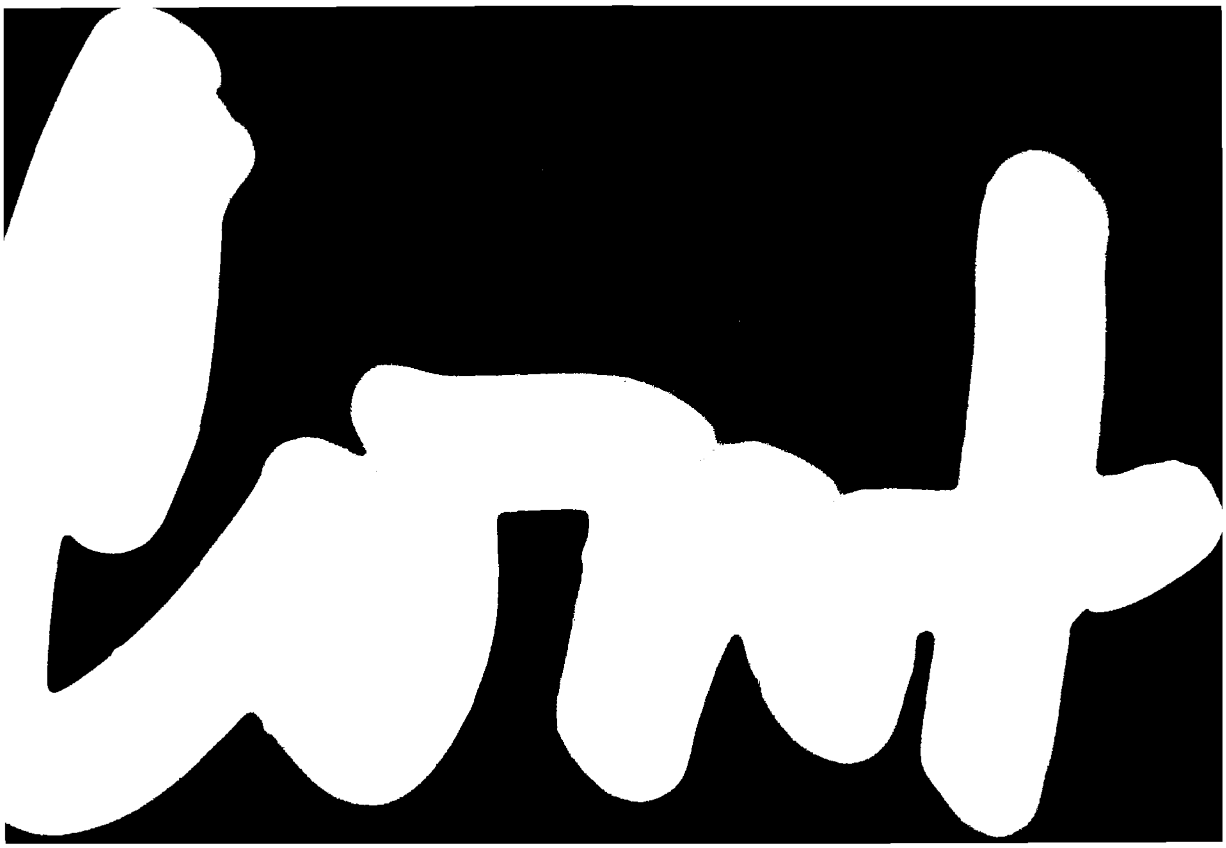
Major Frank Yohannan, USMC
Headquarters, Marine Corps
(Code MPI-20)
Washington, DC 20380

END

FILMED

6-85

DTIC

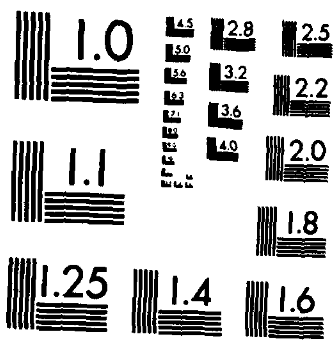


AD-A154 184 BAYESIAN FACTOR ANALYSIS(U) IOWA UNIV IOWA CITY
S I MAYEKAWA 23 MAR 85 TR-85-3-ONR N00014-83-C-0514

3/3

UNCLASSIFIED F/G 12/1 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

SUPPLEMENTARY

INFORMATION

June 21, 1985

BAYESIAN FACTOR ANALYSIS

Shin-ichi Mayekawa
ONR Technical Report 85-3

CORRECTIONS

Page	Line	Correction
Abstract		Replace with enclosed revised abstract.
13	18	(2.2.23) $A = D^{\frac{1}{2}}Q(L + Ir)^{\frac{1}{2}}$ should be (2.2.23) $A = D^{\frac{1}{2}}Q(L - Ir)^{\frac{1}{2}}$
51	12	$= \lvert AC_FA' + D\rvert^{(-\frac{1}{2})N}\ \mathrm{Exp}((-\frac{1}{2})\ \mathrm{tr}\ (Y'YD^{-1}))$ should be $= \lvert AC_FA' + D\rvert^{(-\frac{1}{2})N}\ \mathrm{Exp}((-\frac{1}{2})\ \mathrm{tr}\ (Y'Y(AC_FA' + D)^{-1}))$
54	12	given $\underline{v}$ should be given $\underline{u}$
58 and 60	10 and 7&8	change $\Sigma_{i=1}^{N}$ to $\Sigma_{i=1}^{n}$
60	9	$n2\ \mathrm{tr}\ W^{22}V^* + \text{const.}$ should be $n2\ \mathrm{tr}\ W^{22}V^* + n\ \ln\lvert W\rvert + \text{const.}$
59	19	the line should read: where $C = [X', Y^{*\prime}]\ J\ [X', Y^{*\prime}]'$, the mean corrected SSCP
77	15	$d\underline{a}_j$ should be $f(\underline{a}_j \mid d_j, H, S, \underline{z}_{(j)})\,d\underline{a}_j$
78	13	$E_F(A,D,H \mid S,Y)$ should be $E_{AF}(D,H \mid S,Y)$
91	20	$u_j = RSS_j^{**} + \mathrm{tr}\ (F^{*\prime}F^* + NV^*)\ Q_j + \underline{a}_j^{*\prime}V^*\underline{a}_j^*$ should be $u_j = RSS_j^{**} + \mathrm{tr}\ (F^{*\prime}F^* + NV^*)\ Q_j + N\underline{a}_j^{*\prime}V^*\underline{a}_j^*$
94	18	$A'A$ should be AA'
140	5(f)	the (4.4) element is .07
153	6(f) 6(g)	delete Factor Loadings and Factor Scoring Weights
154	6(h)	delete Factor Loadings and Factor Scoring Weights

Bayesian Factor Analysis*

Shin-ichi Mayekawa

The University of Iowa

Abstract

A new Bayesian procedure for factor analysis is developed in which factor scores as well as factor loadings and error variances are treated as parameters of interest. The presentation is fully Bayesian in the sense that all the parameters have prior distributions and the posterior mode of a subset of the parameters is used as the point estimate.

The model is a standard one where the observations are expresssed as the sum of the linear combination of factor scores, with factor loadings being the weights, and a normal error term. As the prior distribution the following exchangeable form is assumed:

A factor score vector for each observation has a common normal distribution.

A factor loading vector for each variable has a common normal distribution.

A error variance for each variable has a common inverted chi square distribution.

When the exchangeability of all the observations/variables is in question observations/variables may be divided into several subsets and the observations/variables within each subset may be treated as exchangeable.

Since the posterior marginal distribution of factor loadings and error variances can be expressed as the product of the covariance-based likelikhood and the prior distributions of factor loadings and error variances the proposed method includes both the random and the fixed factor analysis models.

The mode of the hyperparameters is first derived from their posterior marginal distributions and conditional on those values the mode of error variance is derived from their posterior marginal distributions. Then, conditional of those estimates, the point estimate of factor scores and factor loadings are derived as the joint or the marginal mode of the posterior distribution of factor scores and factor loadings depending on the investigator's interest.

The marginalization is done via some variations of the EM algorithm and it is found that the different variations result in almost identical estimates. It is also found that the effect of the prior distribution of error variances is such that it reduces the number of local maxima. Finally, by specifying a priori zeros in the locational hyperparameters of factor loadings, a simple structure can be obtained without rotation.

*Support for this research was provided under contract #N00014-83-C-0514 with the Personnel Training Branch of the Office of Naval Research, Melvin R. Novick, Principal Investigator. I am indebted to Professor Novick and Dr. Ming-mei Wang for their comments on earlier drafts. Also, I would like to thank Professor Tom Leonard of the University of Wisconsin. Some of the methods used in chapter IV were originally proposed by him.

END

FILMED

9-85

DTIC